Aalto University
School of Science
Degree Programme in Engineering Physics and Mathematics

Niko Lietzén

New Approach to Complex Valued ICA:
From FOBI to AMUSE

Master's Thesis
Espoo, March 22, 2016

Supervisor: Assistant professor Pauliina Ilmonen, Aalto University
Advisor: Assistant professor Pauliina Ilmonen

Aalto University
School of Science
Degree Programme in Engineering Physics and Mathematics

ABSTRACT OF
MASTER'S THESIS

Author: Niko Lietzén

Title:
New Approach to Complex Valued ICA: From FOBI to AMUSE

Date: March 22, 2016 **Pages:** vii + 78

Major: Systems and Operations Research **Code:** F3008

Supervisor: Assistant professor Pauliina Ilmonen

Advisor: Assistant professor Pauliina Ilmonen

In this Master's Thesis, we study Independent Component Analysis (ICA) for complex valued random variables. In ICA, we assume that the elements of an observed p variate random vector are linear combinations of an unobservable p variate vector with mutually independent components. The goal is to estimate an unmixing matrix that transforms the observed p variate vector to the independent components. ICA has numerous applications and recently the applications have been studied in various fields of science. In this Thesis, we consider ICA for independent and identically distributed observations and for time series data.

We present a new approach for complex valued time series ICA. We have formulated a complex version of the AMUSE (Algorithm for Multiple Unknown Signals Extraction) transformation. We compare the AMUSE transformation to the FOBI (Fourth-Order Blind Identification) transformation. Moreover, we compare these ICA transformations to Principal Component Analysis (PCA) and Invariant Coordinate Selection (ICS). We consider theoretical properties of AMUSE and FOBI transformations and we conduct a simulation study to demonstrate that FOBI and AMUSE work in different settings. We have derived a complex version of the minimum distance index to perform the comparison. The most promising application for our method is the functional magnetic resonance imaging, where the collected data is complex valued time series. We also have a short example related to fMRI.

We demonstrate our method with an image data example. In the example, we have signals that form two-dimensional fractals. We then mix the fractals and apply AMUSE transformation to the mixed images. The effectiveness of our method is then easy to verify, since the unmixed images are almost identical to the original images.

Keywords: complex valued random variable, independent component analysis, location functional, multivariate analysis, performance index, scatter functional, time series

Language: English

Aalto-yliopisto
Perustieteiden korkeakoulu
Teknillisen fysiikan ja matematiikan koulutusohjelma

DIPLOMITYÖN
TIIVISTELMÄ

Tekijä: Niko Lietzén

Työn nimi:
Uusia näkökulmia kompleksiarvoiseen riippumattomien komponenttien analyysiin

Päiväys:	22. Maaliskuuta 2016	**Sivumäärä:**	vii + 78
Pääaine:	Systeemi- ja operaatiotutkimus	**Koodi:**	F3008
Valvoja:	Apulaisprofessori Pauliina Ilmonen		
Ohjaaja:	Apulaisprofessori Pauliina Ilmonen		

Tämän diplomityön tarkoituksena on tutkia riippumattomien komponenttien analyysia kompleksiarvoisille satunnaismuuttujille. Riippumattomien komponenttien analyysissa oletetaan, että havaitun p ulotteisen vektorin alkiot ovat lineaarikombinaatioita havaitsemattomasta p ulotteisesta vektorista, jolla on toisistaan riippumattomat komponentit. Tavoitteena on löytää sekoitusmatriisin käänteismatriisi, jonka avulla havaittu p ulotteinen vektori voidaan muuntaa riippumattomiksi komponenteiksi. Johtuen riippumattomien komponenttien analyysin lukuisista sovelluskohteista, sitä tutkitaan aktiivisesti niin sovellusten kuin teoriankin näkökulmasta. Tässä työssä tarkastelemme riippumattomien komponenttien mallia, jossa havainnot ovat riippumattomia ja samoin jakautuneita. Lisäksi tarkastelemme riippumattomien komponenttien aikasarjamalleja.

Työssä esitellään uusi lähestymistapa kompleksiarvoisien aikasarjojen riippumattomien komponenttien löytämiseksi. Olemme määritelleet kompleksisen version AMUSE (Algorithm for Multiple Unknown Signals Extraction) menetelmästä. Vertaamme AMUSE transformaatiota FOBI (Fourth-Order Blind Identification) transformaatioon. Lisäksi vertaamme näitä riippumattomien komponenttien analyysin menetelmiä pääkomponenttianalyysiin ja invarianttien koordinaattien valintaan. Tarkastelemme AMUSE ja FOBI menetelmien teoreettisia ominaisuuksia ja havainnollistamme eroja simulaatiotutkimuksella. Työssä on johdettu vertailun suorittamiseksi kompleksinen versio lyhimmän etäisyyden indeksistä. Lupaavin sovelluskohde menetelmällemme on funktionaalinen magneettikuvaus, jossa havaittu aineisto on kompleksiarvoista aikasarjaa.

Havainnollistamme menetelmämme toimivuutta esimerkillä, jossa kompleksiarvoiset signaalit muodostavat piirrettynä kaksiulotteisia fraktaaleja. Sekoitamme alkuperäiset fraktaalit ja käytämme AMUSE transformaatiota sekoitettuihin kuviin. Löydetyt kuvat ovat lähes identtisiä alkuperäisten kuvien kanssa.

Asiasanat:	aikasarja, hajontafunktionaali, kompleksiarvoinen satunnaismuuttuja, lokaatiofunktionaali, monimuuttujamenetelmät, riippumattomien komponenttien analyysi, suorituskyvyn mittari
Kieli:	englanti

Acknowledgements

This study was carried out at the Department of Mathematics and Systems Analysis of the Aalto University School of Science. I would like to thank the instructor and supervisor of this thesis, Professor Pauliina Ilmonen for her excellent guidance during this project. I thank Pauliina for introducing me to this interesting research field and for always having time to answer my questions. Her expertise and insights made every part of this project go very smoothly. I would also like to thank everyone working in the Department of Mathematics and Systems Analysis for creating a nice and creative working environment.

I would also like to thank my parents, friends and especially Katariina for supporting me in everyday things and giving me the opportunity to spend countless late nights working on my research.

Helsinki, March 22, 2016

Niko Lietzén

Abbreviations and Acronyms

AMUSE	Algorithm for Multiple Unknown Signals Extraction
BSS	Blind Source separation
cdf	cumulative distribution function
FOBI	Fourth-order Blind Identification
ICA	Independent Component Analysis
ICS	Invariant Coordinate Selection
i.i.d.	independent and identically distributed
MD index	Minimum Distance index
PCA	Principal Component Analysis
pdf	probability density function
ACov_τ	Autocovariance matrix with lag τ
Cov_4	Scatter matrix based on fourth moments
$A \leftarrow B$	A is replaced by B
A^*	Conjugate transpose of A
D	diagonal matrix
J	sign change matrix
L	phase shift matrix
P	permutation matrix

Contents

Chapter 1

Introduction

In the modern information society there are countless means of measuring and acquiring vast amounts of data. Currently, we have more data that we can process. Separating noise from the underlying and interesting structures that hide beneath the surface is an important focus of research in various fields of science. Finding these unobservable and hidden structures is sometimes crucial in exposing the true nature of the data. Furthermore, in many applications the phenomenon in interest is not directly measurable, but mixtures of something measurable.

The problem described above is called the Blind Source Separation (BSS) problem. The word blind is used for denoting all identification or inversion methods based on output observations only. In the problem of BSS, we assume that an unknown linear system consists of several inputs and outputs. The aim is to recover the input vectors by estimating an unmixing matrix. The number of phenomenon where we cannot directly measure the interesting source signals is vast and include for example electroencephalography (EEG), telecommunications signals, wind prediction, financial time series, digital image processing, just to mention a few. See Anemüller et al. (2003); Adali et al. (2011); Mandic et al. (2009); Kiviluoto and Oja (1998); Comon and Jutten (2010).

According to Comon and Jutten (2010), the source separation problem was originally formulated in the 1980s in the framework of neural modeling (Roll (1981)) and independently in the framework of communications (Bar-Ness et al. (1982)). Immediately the papers contributed in the mid-1980s drew the attention of the signal processing community. Since then the blind source separation (BSS) problem has been addressed widely in literature. Several algorithms have been proposed to solve the BSS problem. However, the study on the statistical properties has been insufficient, especially when we face complex valued data. Recently, the interest towards BSS has been

increasing, see Chen and Bickel (2006); Ilmonen et al. (2011); Matteson and Tsay (2011); Samworth et al. (2012); Nordhausen (2014) and their bibliographies. Theoretical analysis of some BSS functional has been studied in e.g Ilmonen et al. (2010a); Nordhausen et al. (2011b) and Ollila (2010).

A particularly interesting submodel of BSS is the Independent Component Analysis (ICA) model, where we further assume that the components of the unobservable source vector are mutually independent. ICA was originally based on the use of the regular covariance matrix and the scatter matrix based on fourth moments, see Cardoso (1989). The transformation was later named as the FOBI (Fourth-Order Blind Identification) transformation. By assuming statistical independence, we can identify source signatures without any a priori model for propagation or reception, provided that the source vectors are statistically independent and at most one them is normally distributed. For more, see Hyvärinen et al. (2001) and Comon (1994).

In this Thesis, we consider real and complex valued ICA based on simultaneous use of two scatter matrix functionals. The use of two scatter matrices has been considered in Nordhausen et al. (2008); Oja et al. (2006) (real data) and Ollila et al. (2008)(complex data). The asymptotical properties have been studied in Ilmonen et al. (2010a, 2012b)(real data) and Ilmonen (2013)(complex data).

However, the papers above do not consider the fact, that there is often a temporal (e.g. time series) or spatial (e.g. image analysis) dependence between the observations. This is in contrast to basic ICA model where the observations have no particular order. In many applications, the mixed signals are not random variables but time signals or time series. Ignoring the time series form can lead to loss of crucial information. In time series context a scatter matrix type approach is obtained using simultaneously two autocovariance matrix functionals with different lags. This approach is called AMUSE (Algorithm for Multiple Unknown Signals Extraction). ICA based on autocovariance matrix functionals has been applied to second-order stationary time series for real valued data, see Tong et al. (1990). The asymptotic properties of the AMUSE transformation under general conditions has been considered in Miettinen et al. (2012).

The complex valued AMUSE has not previously been studied in the literature. In this thesis we introduce the complex version of the AMUSE transformation. The applications, where the complex valued AMUSE can be utilized include functional MRI (Adali and Calhoun (2007)), prediction of wind forecast profile (Mandic et al. (2009)), signal processing (Adali et al. (2011)), spatiotemporal analysis of EEG (Anemüller et al. (2003)) and many more.

We compare the complex AMUSE to the complex version of FOBI. The

FOBI transformation and asymptotic results have been studied in both the real and the complex case, see Ilmonen et al. (2012b) and Ilmonen (2013). The FOBI transformation does not take into account the time dependence of the observations. For the comparison, we have derived complex version of the Minimum Distance (MD) index. The MD index is a tool to measure performance of different methods in simulation studies.

This Thesis is organized as follows. In Chapter 2 we give background to the independent component problem and discuss the required functionals for the FOBI and AMUSE transformations. Furthermore, some basic statistical concepts are reviewed and we give geometrical interpretation for some complex valued statistics. Additionally, complex numbers are reviewed in the last Section of Chapter 2. In Chapter 3 we review the concept of an invariant coordinate system and the whitening and the principal component transformations. In Chapter 4 we introduce the general framework of ICA using two scatter matrices and examine the AMUSE and FOBI transformations. Additionally, we introduce the complex version of the Minimum Distance (MD) index which is a performance measure for separation functionals. In Chapter 5, we have complex valued simulated examples and we measure the performance of FOBI and AMUSE using the MD index. In Chapter 6 we have a complex image source separation example involving separation of fractal images and an example related to functional magnetic resonance imaging. Finally, in Chapter 7 we introduce some possible future research related to the topic. In Appendix A, we have derived some necessary formulas. In Appendix B, we present some proofs for our theorems. In Appendix C, we give the required R codes to replicate our results, since the majority of the existing R packages do not support complex numbers.

Chapter 2

Background and Definitions

A persistent problem in statistics and related fields of science is the trouble of finding a transformation for data such that the essential structures are made more visible and accessible. Assume that the data consists of p observed variables. Furthermore, we have n observations related to every variable. We can then denote our data by x_i, where every x_i is a p variate random vector and $i = 1, \ldots, n$. Depending on the context, i can be interpreted as time if the ordering of the observations matters. We then assume a linear model $x_i = \Omega z_i + \mu$, where x_i is observed and we have little information on Ω and z_i. The vector z_i is some k variate random vector and Ω is a $p \times k$ matrix. In this Thesis, we restrict to the case $k = p$. Our goal is to find a transformation for the data such that the latent z_i becomes visible. Usually, only linear transformations are considered, since the computation is faster and interpretation of the results is far more simple in the linear case.

The problem can also be formulated the following way. We have an original p variate source signal that enters a linear mixing system. We can only observe a p variate signal coming out of the system. The original signal and the mixing system are both unknown. Our aim is to reverse the effects of the mixing system by finding an unmixing matrix. This is the BSS (Blind Source Separation) problem. The word blind refers to the fact that the original source is unknown to us.

A classical example is the cocktail-party problem. Consider a situation where p individuals are speaking in the same room. You have p microphones in different locations. The microphones give you p different speech signals that are mixtures of the general noise in the room. The aim is to find estimates for the original speech signals. In most cases, it is realistic to assume that the p speech signals are statistically independent. We can then reduce the problem to the IC (Independent Component) problem. Actually, Independent Component Analysis (ICA) was originally developed to deal with

problems that are closely related to the cocktail-party problem. Since then the applications of ICA have grown substantially, e.g. applications in electroencephalogram (EEG), econometrics and functional Magnetic resonance imaging (fMRI), just to mention a few. See Comon and Jutten (2010) for a collection of applications.

In this Chapter we review the definitions of some basic building blocks in Independent Component Analysis (ICA), starting with the proper definition of statistical independence. We give the definitions of location and scatter functionals and give some well known examples of them. Furthermore, we briefly discuss estimation techniques in general and give the sample versions of the estimators that we use in this Thesis. We simultaneously consider estimators suited for independent and identically distributed (i.i.d.) and time series random vectors. All the variables and estimators are considered in the multivariate setting. We first give the definitions assuming that the variables are real valued. We then extend the definitions for complex valued variables. Additionally, we briefly discuss the geometrical interpretation of the estimators in the complex case. The interpretation for e.g. complex correlation is rarely considered in literature.

2.1 Statistical Independence

A key concept in independent component analysis is the definition of statistical independence. In the most simple case, consider two random variables x and y that are mutually independent. This implies that the value of y gives no information on the value of x. For example, let x and y be random signals originating from different physical processes that are in no way related to each other. Such independent processes can be e.g. random variables that are the value of a dice thrown and the value of a coin tossed. Statistical independence is formally defined in terms of cumulative distribution functions.

Definition 2.1.1 (Statistical independence)**.** *Let $F_x(x)$ and $F_y(y)$ denote the cumulative distribution functions of random variables x and y, respectively, and let $F_{x,y}(x, y)$ denote their joint cumulative distribution function. Variables x and y are statistically independent, if*

$$F_{x,y}(x, y) = F_x(x)F_y(y).$$

Assuming that the probability density functions of x and y exists, statistical independence can equivalently be defined by replacing the cumulative distribution functions (cdf) by the corresponding probability density functions (pdf) in Definition 2.1.1, respectively.

Corollary 2.1.1. *Independent random variables x and y satisfy the basic property*

$$\mathbb{E}\left(g(x)h(y)\right) = \mathbb{E}\left(g(x)\right)\mathbb{E}\left(h(y)\right),$$

where $g(x)$ and $h(y)$ are any absolute integrable functions of x and y.

Definition 2.1.1 can be generalized for more than two random variables and random vectors. The result can be found in any premilinary mathematical statistics course book, see e.g. Hogg et al. (2005).

2.2 Location and Scatter Functionals

There are countless of different location and scatter estimators defined in the literature. Different settings require estimators with different statistical properties. For example robustness, convergence, computation time, limiting distributions and efficiency are statistical properties that vary between different estimators. Generally, different estimators do not estimate the same population quantities. The most suitable estimator depends on the data at hand.

We should define our location and scatter functionals in such a way that they change logically whenever the coordinate system is altered. The attribute that guarantees that no coordinate system change affects the interpretation of the results is called equivariance in the context of location and scatter functionals. For example, we want to acquire the same results whether we measure the observations in the metric system or the imperial system.

Consider the following for the definitions in this Section: Let x be a real valued p variate random vector with a cumulative distribution function F_x and let $X = [x_1, ..., x_n]$, where $x_1, ..., x_n$ are i.i.d. observations from the distribution F_x.

Definition 2.2.1 (Location functional)**.** *A p variate vector valued functional $T(F_x)$ is a location functional if it is affine equivariant in the sense that*

$$T(F_{Ax+b}) = AT(F_x) + b,$$

for all full rank $p \times p$ matrices A and for all p variate vectors b.

Definition 2.2.2 (Scatter functional)**.** *A $p \times p$ matrix valued functional $S(F_x)$ is a scatter functional if it is positive definite and affine equivariant in the sense that*

$$S(F_{Ax+b}) = AS(F_x)A^T,$$

for all full rank $p \times p$ matrices A and for all p variate vectors b.

Scatter functionals are required to be positive (semi)definite since a functional that gives negative measures of scatter is not practical in statistical analysis. Let X denote a $p \times n$ data matrix of observations from F_x. Sample location and scatter statistics are obtained when the functionals are applied to the empirical cumulative distribution function, denoted by F_n, of the sample X. The location and scatter sample statistics then satisfy

$$T\left(AX + b1_n^T\right) = AT(X) + b, \tag{2.1}$$

and

$$S\left(AX + b1_n^T\right) = AS(X)A^T, \tag{2.2}$$

for all nonsingular $p \times p$ matrices A and for all p variate vectors b. We use the notation $T(F_n)$ and $S(F_n)$ or $T(X)$ and $S(X)$ for sample statistics. Scatter matrix functionals are often standardized in a way that for the standard normal distribution $S(F_x) = I$. Well known examples of a location and a scatter functional are the expected value and the (population) covariance matrix:

$$T(F_x) = \mathbb{E}(x), \tag{2.3}$$

$$S_1(F_x) = \mathrm{Cov}(F_x) = \mathbb{E}\left((x - \mathbb{E}(x))\,(x - \mathbb{E}(x))^T\right). \tag{2.4}$$

An interesting functional that is closely related to the covariance matrix is the correlation matrix. Note, that the correlation matrix is not a scatter functional since it is not affine equivariant in the required sense. In this Thesis we refer to the matrix of Pearson product-moment correlation coefficients between each of the components in the random vector x, when we speak of correlation matrix. Pearson's product moment correlation matrix is simply a scaled version of the covariance matrix. We can define the correlation matrix the following way:

$$\mathrm{Corr}(F_x) = (\mathrm{diag}(\mathrm{Cov}(F_x))^{-1/2}\,\mathrm{Cov}(F_x)\,(\mathrm{diag}(\mathrm{Cov}(F_x))^{-1/2}, \tag{2.5}$$

if none of the components of x are degenerate and none of the components is fully linearly dependent of the other components. Each diagonal element in the correlation matrix is equal to one and the off-diagonal elements are between -1 and 1. The closer an off-diagonal value is to either -1 or 1, the stronger the linear dependency is between the corresponding components. The components of x are said to be uncorrelated if the covariance matrix $\mathrm{Cov}(F_x)$ is a diagonal matrix. Note that Corollary 2.1.1 reveals that statistical independence is a much stronger property than uncorrelatedness. In the

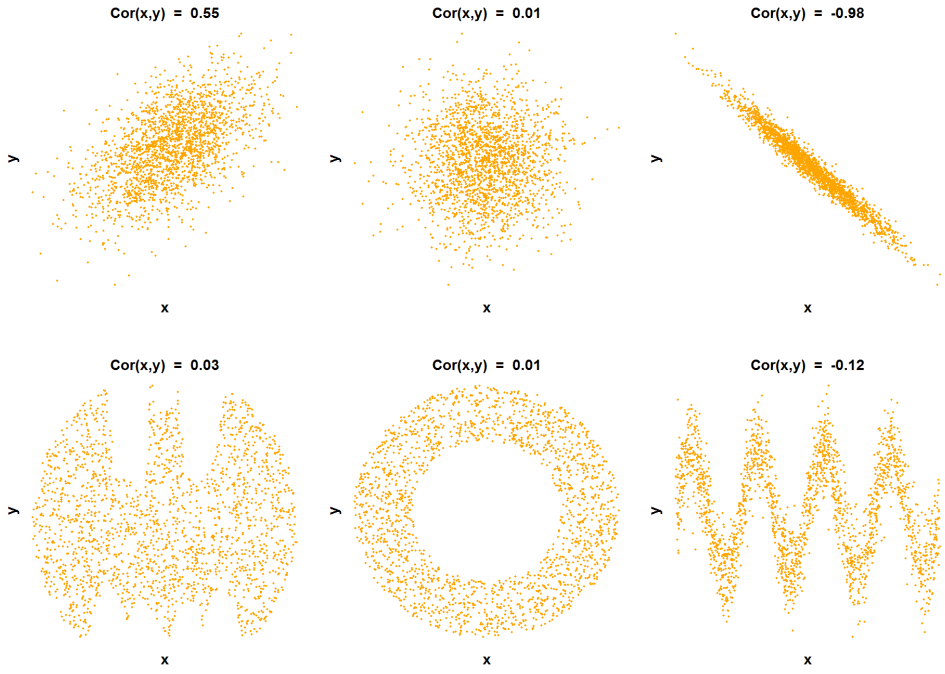


Figure 2.1: Sample correlation of different bivariate distributions.

last row of Figure 2.1 we see bivariate random variables that have small sample correlations. However, it is clear that these variables are not independent. A special case is normally distributed random variables. Uncorrelatedness and independence are equivalent in the case of normal random variables. This is a special property of the multivariate normal distribution. Hereby, we require alternative ways to measure scatter alongside covariance. One alternative measure that we use in this Thesis is the scatter matrix based on fourth moments:

$$S_2(F_x) = \mathrm{Cov}_4(F_x) = \\ \frac{1}{p+2}\mathbb{E}\left((x-\mathbb{E}(x))\,(x-\mathbb{E}(x))^T\,S_1(F_x)^{-1}\,(x-\mathbb{E}(x))\,(x-\mathbb{E}(x))^T\right), \tag{2.6}$$

where $S_1(F_x)$ is the covariance matrix. The scatter matrix based on fourth moments can be used to measure multivariate kurtosis. Under symmetry assumption, kurtosis can be seen as a measure of the peakedness (or the strength of the shoulders), or of the tail thickness of a probability distribution. The functional $\mathrm{Cov}_4(\cdot)$ is scaled such that in the case of standard multivariate normal distribution $\mathrm{Cov}_4(F_x) = I_p$. The scatter matrix based on fourth moments is not defined for distributions that do not have theoretical fourth moments.

The off-diagonal elements in $\mathrm{Cov}_4(\cdot)$ can be used to measure cokurtosis. Random variables that satisfy Definition 2.1.1 of statistical independence have a diagonal scatter matrix based on fourth moments. In finance, cokurtosis is sometimes used as a supplement to the covariance in risk estimation, see Hwang and Satchell (1999). Figure 2.2 contains the probability density functions of some well known univariate distributions. We formed random vectors that have independent components following the different univariate distributions and calculated the theoretical $\mathrm{Cov}_4(\cdot)$ matrices. The results are in Table 2.1. Note that, p has an effect on the diagonal elements on all other distributions except the standard normal distribution. Let x be a p variate random vector with finite fourth moments and independent components. We now have

$$
\begin{aligned}
&\frac{1}{p+2}\mathbb{E}\left((x-\mathbb{E}(x))(x-\mathbb{E}(x))^T \mathrm{Cov}(F_x)^{-1}(x-\mathbb{E}(x))(x-\mathbb{E}(x))^T\right)\\
&=\frac{1}{p+2}\mathbb{E}\left(yy^T\mathrm{Cov}(F_x)^{-1}yy^T\right)\\
&=\frac{1}{p+2}\mathbb{E}\left(\begin{pmatrix}y_1\\ \vdots\\ y_p\end{pmatrix}\begin{pmatrix}y_1 & \dots & y_p\end{pmatrix}\begin{pmatrix}d_1^{-1} & \dots & 0\\ \vdots & \ddots & \vdots\\ 0 & \dots & d_p^{-1}\end{pmatrix}\begin{pmatrix}y_1\\ \vdots\\ y_p\end{pmatrix}\begin{pmatrix}y_1 & \dots & y_p\end{pmatrix}\right)\\
&=\frac{1}{p+2}\mathbb{E}\left(\begin{pmatrix}\frac{y_1^4}{d_1}+y_1^2\sum_{i\neq 1}^{p}\frac{y_i^2}{d_i} & \dots & \cdot\\ \vdots & \ddots & \vdots\\ \cdot & \dots & \frac{y_p^4}{d_1}+y_p^2\sum_{i\neq p}^{p}\frac{y_i^2}{d_i}\end{pmatrix}\right)\\
&=\frac{1}{p+2}\begin{pmatrix}\frac{\mathbb{E}\left(y_1^4\right)}{\mathbb{E}\left(y_1^2\right)}+(p-1)\mathbb{E}\left(y_1^2\right) & \dots & 0\\ \vdots & \ddots & \vdots\\ 0 & \dots & \frac{\mathbb{E}\left(y_p^4\right)}{\mathbb{E}\left(y_p^2\right)}+(p-1)\mathbb{E}\left(y_p^2\right)\end{pmatrix},
\end{aligned}
$$

where $\mathrm{Cov}(F_x)=\mathrm{diag}(d_1,\dots,d_p)=\mathrm{diag}\left(\mathbb{E}(y_1^2),\dots,\mathbb{E}\left(y_p^2\right)\right)$ and $y=(x-\mathbb{E}(x))$.

Asymptotically, if the dimension $p\to\infty$, we have

$$
\xrightarrow{p\to\infty}\begin{pmatrix}\mathbb{E}(y_1^2) & \dots & 0\\ \vdots & \ddots & \vdots\\ 0 & \dots & \mathbb{E}\left(y_p^2\right)\end{pmatrix}=\mathrm{Cov}(F_x).
$$

Note that if y_i and y_j are independent, then $i\neq j\to\mathbb{E}(y_iy_j)=\mathbb{E}(y_i)\mathbb{E}(y_j)$. The required central moments are derived in Appendix A.1. When p approaches infinity, we get the regular covariance matrix. However, in this

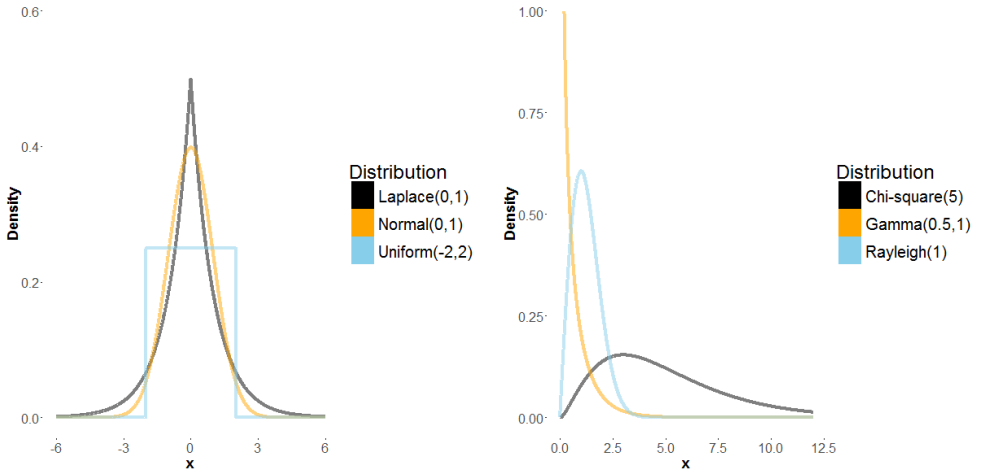


Figure 2.2: The probability density functions of some symmetric and asymmetric distributions.

Table 2.1: The diagonal elements of $\mathrm{Cov}_4(\cdot)$ for different p variate distributions.

Distribution	$p=1$	$p=3$	$p=\infty$
$\mathcal{N}(0,1)$	1	1	1
Laplace$(0,1)$	4	3.2	2
Unif$(-2,2)$	0.8	1.01	1.33
$\chi^2(5)$	18	14.8	10
Gamma$(0.5,1)$	12	8.8	4
Rayleigh(1)	0.46	0.45	0.43

Thesis we do not consider infinite dimensional data. We apply the $\mathrm{Cov}_4(\cdot)$ only to finite dimensional whitened data, see Section 4.1.

When we are dealing with time series observations, there is usually at least some serial dependence between the observations. This is different for i.i.d. data, where there is no dependence between the observations. We often acquire more information related to the structure of the variable of interest when we take the time dependence into account. Now assume that there is a spatial or temporal dependency in x and this is highlighted by the notation x_t. Let x_t be a real valued p variate random vector with a cumulative distribution function F_{x_t}, $t \in \mathbb{T}$ and let $X = [x_1, ..., x_n]$, where $x_1, ..., x_n$ is time series data from the distribution F_{x_t}. A common scatter functional that utilizes the dependency structure of the data is the autocovariance matrix

functional with a chosen τ:

$$S_\tau(F_{x_t}) = \mathrm{ACov}_\tau\left(F_{x_t}\right) = \mathbb{E}\left(\left(x_r - \mathbb{E}(x_r)\right)\left(x_s - \mathbb{E}(x_s)^T\right)\right), \quad r, s \in \mathbb{T}, \quad (2.7)$$

where $s - r = \tau$ is the lag of the autocovariance matrix functional. Note that for stationary processes the autocovariance matrix functional is equivalent for τ and $-\tau$. Stationarity is defined formally in Section 2.4. Furthermore, we get the regular covariance matrix if we set $\tau = 0$. Note that the autocovariance matrix functional is not necessarily symmetric when $\tau \neq 0$ and hereby it is not a scatter matrix functional. We can symmetrize the autocovariance matrix the following way

$$S_\tau\left(F_{x_t}\right)_{\mathrm{symm}} = \frac{1}{2}\left(S_\tau(F_{x_t}) + S_\tau(F_{x_t})^T\right),$$

where $S_\tau\left(F_{x_t}\right)_{\mathrm{symm}}$ is a scatter matrix functional for stationary processes.

There are several other location and scatter functional defined in literature, having different properties (robustness, efficiency, limiting multivariate normality, computation speed, etc.). See e.g. Davies (1987), Lopuhaä (1989), Kent et al. (1996) and Maronna et al. (2006).

2.3 Independence Property

Several scatter matrix type functionals do not achieve affine equivariance in the sense of Definition 2.2.2. However, in many applications, it is sufficient that the functionals are so called shape functionals. Let x be a real valued p variate random vector with a cumulative distribution function F_x.

Definition 2.3.1 (Shape Functional)**.** *If a positive definite $p \times p$ matrix-valued functional $S(F_x)$ satisfies*

$$S(F_{Ax+b}) \propto AS(F_x)A^T,$$

for all nonsingular $p \times p$ matrices A and for all p variate vectors b, then $S(F_x)$ is called a shape functional.

From Definition 2.3.1 it follows that every scatter matrix functional is also a shape functional. In general, different shape functionals measure different population quantities. However, under the assumption of multivariate ellipticity shape functionals measure the same population quantities up to constant multiplication. To make different shape functionals comparable under ellipticity they are often normalized, such that for example $\mathrm{tr}(S(F_x)) = p$

or $||S(x)|| = 1$, where $||\cdot||$ is some matrix norm. For more details, see Paindaveine (2006). A particularly interesting class of scatter functionals consists of those that have the independence property.

Definition 2.3.2 (Independence Property)**.** *If the scatter functional $S(F_x)$ is a diagonal matrix for all random vectors x that have statistically independent components, then $S(\cdot)$ is said to have the independence property.*

Most scatter functionals found in literature possess the independence property only if all the components of x have symmetric distributions. The regular covariance matrix and the scatter matrix based on fourth moments are examples of scatter matrices with the independence property. However, for any scatter functional $S(\cdot)$, a symmetrizied version can be constructed using the following theorem, see Oja et al. (2006).

Theorem 2.3.1. *Let x_1 and x_2 be independent random vectors with the same cumulative distribution function F_x. Then every scatter and shape functional $S(F_x)$ can be symmetrized by*

$$S_{\text{sym}}(F_x) := S(F_{x_1 - x_2}).$$

The symmetrized scatter or shape matrix functional has the independence property.

2.4 Stationary Stochastic Processes

The time series that interest statisticians are usually stochastic, rather than deterministic. Stationarity is an important concept in time series analysis. A stationary time series is a process that has a constant expected value, variance and autocorrelation over time. Many statistical forecasting methods are based on the assumption that the time series is stationary or that it can be stationarized trough the use of transformations. It is convenient to predict stationary processes since the statistical properties stay the same in the future as they have been in the past. Furthermore, sample statistics such as mean, variance and autocovariance are meaningful only if the time series is stationary. If the process is not stationary, the sample statistics become functions of time. Typical non stationary behavior include trends, cycles and random-walking. The formal definition of stationary is given below.

Definition 2.4.1 (Stationary Process)**.** *The p variate time series $\{x_t, t \in \mathbb{Z}\}$*

is stationary if the following three conditions hold

$$\begin{aligned} &(S1) \qquad \mathbb{E}(F_{x_t}) = \mu, \quad \textit{for all } t \in \mathbb{Z}, \\ &(S2) \qquad \mathrm{Cov}\,(F_{x_t}) = \Sigma, \quad \textit{for all } t \in \mathbb{Z}, \\ &(S3) \qquad \mathrm{ACov}_\tau\,(F_{x_t}) = \Sigma_\tau \quad \textit{for all } t, \tau \in \mathbb{Z}, \end{aligned}$$

where μ, Σ and Σ_τ are finite constants.

Stationarity given in Definition 2.4.1 is usually referred to as weak stationarity, covariance stationarity or second-order stationarity. However, in this Thesis the term stationarity will refer to a process with properties specified by Definition 2.4.1. For more wider definitions of stationarity, see Brockwell and Davis (2013).

2.5 Estimation Theory

An important issue encountered in various fields of science is the estimation of quantities of interest from a given finite set of uncertain measurements. There are several estimation techniques developed for different situations. For example, we require different estimators for constant and time-varying variables. Finding the appropriate estimators is an import step in the construction of a mathematical model that fits the data.

Let x be a random variable with a cumulative distribution function $F_x(\theta)$, where $\theta = (\theta_1, \ldots, \theta_m)$ is an unknown parameter vector. Estimation methods can be divided into two classes, depending on whether the parameter θ is assumed to be a deterministic constant or a random variable. The estimation in the latter case often involves assumptions related to the probability distribution function of θ and are usually called Bayesian estimation methods. The name of the Bayesian methods come from the well-known Bayes' rule, which is utilized in Bayesian methods. In this Thesis we will not consider Bayesian estimation. We take the frequentist approach and assume that the underlying parameter vector θ is not random.

Let $X = [x_1, \ldots, x_n]$ where $x_1, \ldots, x_n$ are i.i.d. observations from the distribution $F_x(\theta)$. We can estimate the parameter θ by an estimation formula

$$\hat{\theta} = g(X) = g(x_1, \ldots, x_n), \tag{2.8}$$

where $g(\cdot)$ is an estimation function of our choice. There are multiple different paths to perform the estimation. For example, if the distribution is known up to the parameter of interest we can apply the maximum likelihood method, see for example Hogg et al. (2005). In this Thesis we mainly consider unbiased estimators that are consistent.

Definition 2.5.1 (Unbiased Estimator)**.** *An estimator $\hat{\theta}$ is unbiased if*

$$\mathbb{E}(\hat{\theta}) = \theta.$$

Definition 2.5.2 (Convergence in Probability)**.** *We say that an estimator $\hat{\theta}_n$ converges to the true value of the parameter vector θ in probability, if for all $\varepsilon > 0$, as the sample size n grows without a limit:*

$$\lim_{n\to\infty} \mathbb{P}\left(|\hat{\theta}_n - \theta| \geq \varepsilon\right) = 0.$$

We then write $\hat{\theta}_n \xrightarrow[P]{n\to\infty} \theta$.

If an estimator converges in probability, it is called a consistent estimator. Let x be a real valued p variate random vector with a cumulative distribution function F_x and let $X = [x_1, ..., x_n]$, where $x_1, ..., x_n$ are i.i.d. observations from the distribution F_x. The population mean and covariance can be estimated by the well-known formulas:

$$\hat{\mu}_x = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \tag{2.9}$$

$$\hat{S}_1\left(F_n\right) = \hat{S}_1\left(X\right) = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(x_i - \bar{x}\right)^T. \tag{2.10}$$

Furthermore, the scatter matrix based on fourth moments can be estimated using:

$$\begin{aligned}&\hat{S}_2\left(F_n\right) = \hat{S}_2(X) = \\ &\frac{1}{n(p+2)}\sum_{i=1}^{n}\left(\left(x_i - \bar{x}\right)\left(x_i - \bar{x}\right)^T \hat{S}^{-1}(X)\left(x_i - \bar{x}\right)\left(x_i - \bar{x}\right)^T\right).\end{aligned} \tag{2.11}$$

If the second moments of the distribution F_x exists as finite quantities, then the sample covariance matrix is consistent. Under the existence of fourth moments it is also $\sqrt{n}$-consistent and asymptotically normal. For the consistency of the scatter matrix based on fourth moments the existence of fourth moments is required and for $\sqrt{n}$-consistence and asymptotic normality eight moments are required.

Now consider time series observations. Let $X = [x_1, x_2, \ldots, x_n]$ be observations from a real p variate stationary time series x_t with a cdf F_{x_t}. The estimate for the autocovariance matrix functional $S_\tau(F_{x_t})$ is the sample autocovariance matrix:

$$\hat{S}_\tau\left(F_n\right) = \hat{S}_\tau(X) = \frac{1}{n-\tau}\sum_{i=1}^{n-\tau}\left(x_{i+\tau} - \bar{x}\right)\left(x_i - \bar{x}\right)^T, \tag{2.12}$$

where $0 \leq \tau < n$. In literature, see for example Brockwell and Davis (2013), divisor of n is often used rather than $(n - \tau)$. If the second moments of the distribution F_{x_t} exists as finite quantities, then the sample autocovariance matrix is consistent. Under the existence of fourth moments it is also $\sqrt{n}$-consistent and asymptotically normal.

An estimate given by formula 2.12 is not necessarily a symmetric matrix. The symmetrized version that has the scatter sample statistic property can be constructed as

$$\hat{S}_\tau(X)_{\text{symm}} = \frac{1}{2}\left(\hat{S}_\tau(X) + \hat{S}_\tau(X)^T\right), \tag{2.13}$$

where $\hat{S}_\tau(X)_{\text{symm}}$ gives symmetric estimates.

2.6 Independent Component Model

In the independent component model we assume that x is a p variate vector with mutually independent components. There are multiple ways of formulating the IC model. We can permutate or multiply the independent components and the resulting components are also independent. Furthermore, in the complex case we can multiply the independent components with a phase shift matrix and the resulting components are also independent.

Definition 2.6.1 (Independent Component Model)**.** *Let $x = (x_1, x_2, \ldots, x_p)^T$ be a p variate random vector, $z = (z_1, z_2, \ldots, z_p)^T$ be a p variate random vector with mutually independent components and let Ω be a full rank $p \times p$ mixing matrix and μ be a p variate location vector. Independent Component (IC) model is now given by*

$$x = \Omega z + \mu.$$

The assumptions imposed on z define the course of our analysis. Note that z is not observed, only x is directly measurable. In some cases z can have an interpretation and it can be the goal of the analysis to recover it, when only x is observed. Nevertheless, the one of the first challenges we face in practical data analysis is to evaluate the assumptions that we can justify on z.

The following two models are considered in this Thesis and they differ by their assumptions on z.

M1: Independent Component (IC) model. The components of z are independent with $\mathbb{E}(z) = 0_p$, $\text{Cov}(z) = I_p$ and $\text{Cov}_4(z) = \mathcal{D}$, where $\mathcal{D}$ is diagonal with diagonal elements $d_1 > d_2 \ldots > d_p$. Furthermore, at most one of the components of z has a normal distribution.

M2: Time dependent Independent Component (tIC) model. The components of z_t are independent with $\mathbb{E}(z_t) = 0_p$, $\mathrm{Cov}(z_t) = I_p$ and $\mathrm{ACov}_\tau(F_{z_t}) = \Sigma_\tau$ is diagonal for all $\tau = 1, 2, \ldots$.

Note that the assumptions in Model M2 imply the (weak) stationarity and uncorrelatedness of the p variate time series in z_t. The main difference between Models M1 and M2 is that we allow more than one of the components in z to be normally distributed and we assume that z_t has a spatial or temporal dependency between the observations. The normality constraint on the components can be dropped due to the extra information gained from the time structure. For example, time series following some stationary ARMA-model can be normally distributed. However, they have a well defined time structure and hereby we can find the latent z_t vector using this extra information.

Note that the IC model is clearly not uniquely defined. For any $p \times p$ permutation matrix P and any full rank diagonal matrix D, we can always write

$$x = (\Omega PD)\left((PD)^{-1} z\right) + \mu = \tilde{\Omega}\tilde{z} + \mu, \tag{2.14}$$

where $\tilde{z}$ has independent components with expected value zero. Furthermore, if x, z and Ω are complex valued we can write

$$x = (\Omega PDL)\left((PDL)^{-1} z\right) + \mu = \tilde{\Omega}\tilde{z} + \mu, \tag{2.15}$$

where $L = \mathrm{diag}\left(\exp(\theta_1 i), \ldots, \exp(\theta_p i)\right)$ and $\tilde{z}$ has independent components with expected value zero. Hereby, in the real case the IC model is unique up to the order, scales and signs of the independent components. In the complex case, the IC model is unique up to the order, scales and arguments of the independent components. We can solve this identifiability problem by standardizing either z or the mixing matrix Ω. The necessary standardizations of z are given in the definitions of M1 and M2.

2.7 Complex Numbers

In this Section, we review basic properties of complex numbers. Each element, z_j, of a complex valued random vector z has two components, $z_j = x_j + y_j i$, where x_j, $y_j \in \mathbb{R}$ and i is the imaginary unit ($i^2 = -1$). The complex plane is illustrated in Figure 2.3. The angle between the real axis and the vector sum of x and y is often referred to as the argument of the complex number. In some applications the argument is referred to as the

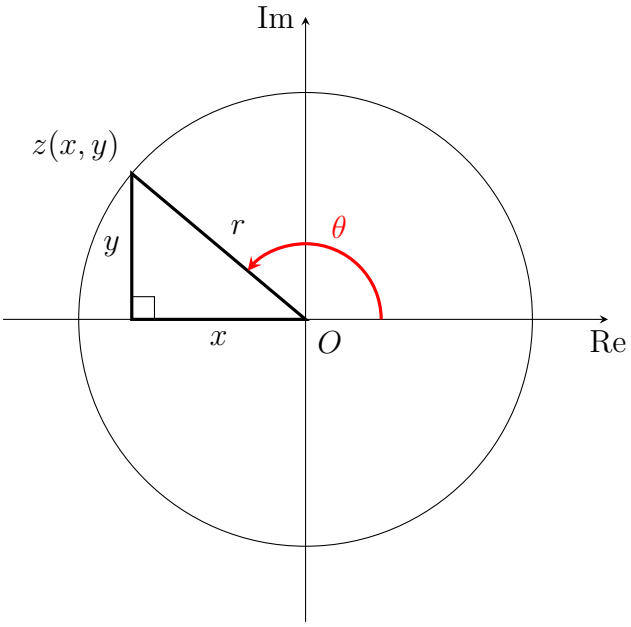


Figure 2.3: An illustration of the complex plane generated using LaTeX package tikzpicture.

phase. The argument is denoted by $\theta \in [0, 2\pi)$. Furthermore, the length of the complex number is often referred to as the modulus and it is denoted by r. The modulus is the length of the vector sum of x and y.

A complex number z_j can also be presented as $z_j = r_j \exp(\theta_j i)$. The complex conjugate of z_j is $\bar{z_j} = x_j - y_j i = r_j \exp(-\theta_j i)$. The cumulative distribution function of z_j is defined as $F(c) = \mathbb{P}(x_j \leq c_1, y_j \leq c_2)$, $c = c_1 + c_2 i$. The expectation of z_j is $\mathbb{E}(z_j) = \mathbb{E}(x_j) + \mathbb{E}(y_j)i$. Note that Definition 2.1.1 of statistical independence holds for complex random variables.

In some situations a complex valued p variate random vector can be transformed to a $2p$ variate real valued random vector. After the transformation from $\mathbb{C}^p$ to $\mathbb{R}^{2p}$, all information is lost related to the dependency between the real and imaginary parts. For a detailed example related to functional MRI, see Adali and Calhoun (2007), where discarding the information related to the argument of the complex numbers leads to suboptimal results.

2.8 Complex Functionals

We can extend many of the real valued functionals to the complex plane by simply replacing the transposes with conjugate transposes. We denote the conjugate transpose of A by A^*. We give the definitions in the same order as

they are introduced in this Chapter. First, the Definition 2.2.1 for location functionals is the same for complex variables as for real variables. However, the definition for a scatter functional is different for complex variables.

Let x be a complex valued p variate random vector with a cumulative distribution function F_x and let $X = [x_1, ..., x_n]$, where $x_1, ..., x_n$ are i.i.d. observations from the distribution F_x.

Definition 2.8.1 (Complex Scatter Functional)**.** *A complex valued $p \times p$ matrix valued functional $S(F_x)$ is a scatter functional if it is positive definite and affine equivariant in the sense that*

$$S(F_{Ax+b}) = AS(F_x)A^*,$$

for all full rank $p \times p$ matrices A and for all p variate vectors b.

The scatter sample statistic then satisfy the following relation:

$$S(AX + b1_n^T) = AS(X)A^*, \tag{2.16}$$

for all nonsingular $p \times p$ matrices A and for all p variate vectors b. Note that A and b can have complex valued elements. The complex valued covariance matrix is defined the following way:

$$S_1(F_x) = \mathbb{E}\left((x - \mathbb{E}(x))\,(x - \mathbb{E}(x))^*\right), \tag{2.17}$$

and the corresponding sample statistic is then

$$\hat{S}_1(X) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})\,(x_i - \bar{x})^*. \tag{2.18}$$

The complex covariance matrix is conjugate symmetric, hereby the diagonal elements are real valued. However, the off-diagonal elements can be complex valued. We can formulate the complex correlation matrix the same way as in the real case using Equation 2.5. The interpretation of complex correlation is not as straightforward as in the real case. Let z_1 and z_2 be complex valued random variables and let σ_{12} be the complex valued correlation between them. Now the modulus of σ_{12} is the measure of linear dependency. The modulus, r, is always between zero and one and we can interpret it similarly to real numbers.

We can see the sign of the correlation from the argument of σ_{12}. Whenever θ is 0 or π, then the correlation matrix contains only real valued elements. The argument is zero or π only if z_1 and z_2 are distributed perpendicularly in the complex plane. Likewise, the σ_{12} has only an imaginary component if

the variables are orthogonally distributed. Otherwise, σ_{12} is a linear combination of a real and an imaginary component. Some complex valued correlations are illustrated in Figure 2.4. Note that for complex random variables, $\mathrm{Cov}(x, y) \neq \mathrm{Cov}(y, x)$. Instead, the relation $\mathrm{Cov}(x, y) = \overline{\mathrm{Cov}(y, x)}$ holds, where $\overline{\mathrm{Cov}(y, x)}$ is the complex conjugate of $\mathrm{Cov}(y, x)$.

The complex scatter matrix on fourth moments is defined the following way:

$$S_2(F_x) = \frac{1}{p+1}\mathbb{E}\left((x - \mathbb{E}(x))(x - \mathbb{E}(x))^* S_1(F_x)^{-1}(x - \mathbb{E}(x))(x - \mathbb{E}(x))^*\right), \tag{2.19}$$

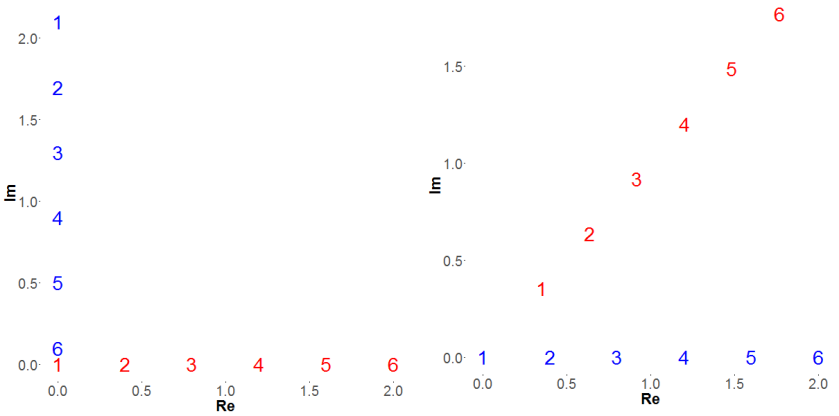
and the corresponding sample statistic is then

$$\hat{S}_2(X) = \frac{1}{n(p+1)}\sum_{i=1}^{n}\left((x_i - \bar{x})(x_i - \bar{x})^T \hat{S}^{-1}(X)(x_i - \bar{x})(x_i - \bar{x})^T\right). \tag{2.20}$$

The interpretation of the complex valued kurtosis is similar to the complex valued correlation. The diagonal elements are again real numbers and hereby the interpretation is exactly the same as for real numbers. The modulus of the off-diagonal elements indicates the strength of the cokurtosis and the argument indicates how the cokurtosis is spread between the real and imaginary axis.

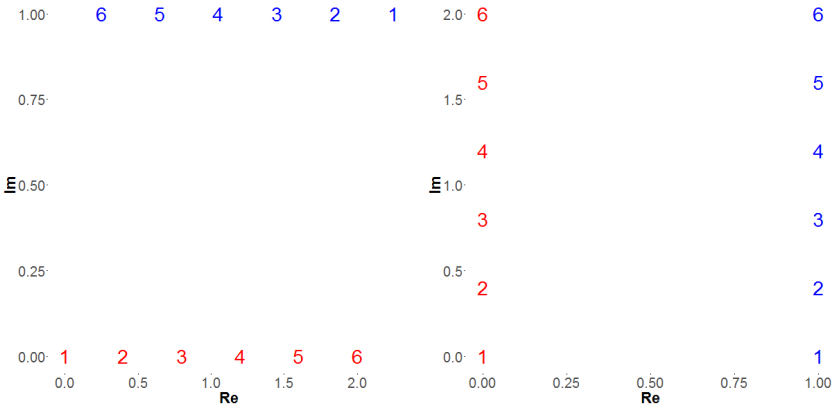
We choose the scaling parameter as $1/(p+1)$ to scale $S_2(\cdot)$ to be 1 for the standard complex normal distribution. The analytical proof is in Appendix A.2. Note that the scaling is different for real valued $S_2(\cdot)$. We also performed a small simulation study to compare the scaling parameters $1/(p+1)$ and $1/(p+2)$ for the complex standard normal distribution. In the simulation we calculated the $S_2(\cdot)$ estimate 1400 times from a p variate random complex standard normal distribution with 10^5 observations. The number of dimensions p was randomly selected from the set $\{3, 4, \ldots, 20\}$ in each iteration. Then we subtracted the identity matrix from the estimate and calculated the Frobenius norm from the subtraction. The results are in Figure 2.5. The histograms show that $1/(p+1)$ is the logical scaling parameter for complex valued random variables. However, the choice of the scaling parameter has no effect in the framework of this Thesis, see Section 4.1. This is due to the whitening part of our transformations.

Now consider complex time series observations. Let $X = [x_1, x_2, \ldots, x_n]$ be observations from a complex p variate stationary time series x_t with a cdf F_{x_t}.



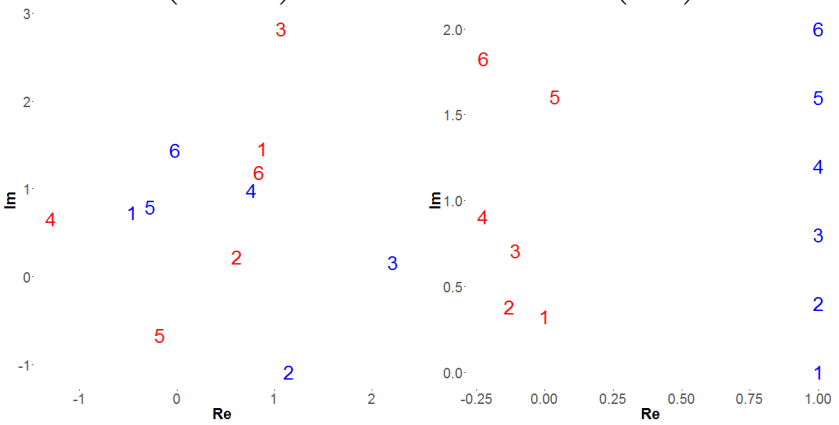
(a) $\begin{pmatrix} 1 & -i \\ i & 1 \end{pmatrix}$

(b) $\begin{pmatrix} 1 & 2^{-\frac{1}{2}}(1-i) \\ 2^{-\frac{1}{2}}(1+i) & 1 \end{pmatrix}$



(c) $\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$

(d) $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$



(e) $\begin{pmatrix} 1 & 0.09 - 0.43i \\ 0.09 + 0.43i & 1 \end{pmatrix}$

(f) $\begin{pmatrix} 1 & 0.95 - 0.06i \\ 0.95 + 0.06i & 1 \end{pmatrix}$

Figure 2.4: Correlation matrix calculated from z, where $z = (z_{\text{red}}, z_{\text{blue}})$. The data points are labeled as numbers, where the numbers denote the corresponding observations.

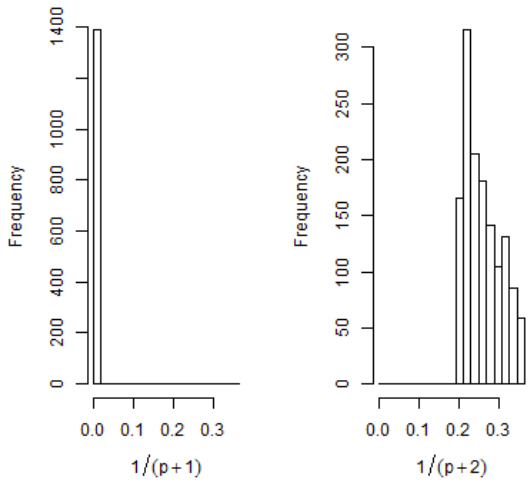


Figure 2.5: Histograms of $||S_2(X) - I_p||_F$ with 1400 iterations using two different scaling factors. The sample X follows the p variate complex standard normal distribution with 10^5 observations and p is randomly selected from $\{3, 4, \ldots, 20\}$ in each iteration.

$$S_\tau\left(F_{x_t}\right) = \mathbb{E}\left(\left(x_r - \mathbb{E}(x_r)\right)\left(x_s - \mathbb{E}(x_s)^*\right)\right), \quad r, s \in \mathbb{T}, \tag{2.21}$$

where $s - r = \tau$. The corresponding sample statistic is then

$$\hat{S}_\tau\left(X\right) = \frac{1}{n-\tau}\sum_{i=1}^{n-\tau}\left(x_{i+\tau} - \bar{x}\right)\left(x_i - \bar{x}\right)^*, \tag{2.22}$$

The symmetrized version is

$$S_\tau(F_{x_t})_{\text{symm}} = \frac{1}{2}\left(S_\tau(F_{x_t}) + S_\tau(F_{x_t})^*\right), \tag{2.23}$$

and the corresponding sample statistic is

$$\hat{S}_\tau(X)_{\text{symm}} = \frac{1}{2}\left(\hat{S}_\tau(X) + \hat{S}_\tau(X)^*\right), \tag{2.24}$$

where $\hat{S}_\tau(X)_{\text{symm}}$ gives conjugate symmetric estimates. Note that the definition of stationarity given in Section 2.4 also holds for complex valued variables.

Chapter 3

Linear Data Transformations

The human eye can simultaneously visualize random variables at maximum in three dimensional spaces. However, in various real life applications, the data we face is more than three dimensional. Fortunately, the information contained in the p variate case can sometimes be summarized by a smaller set of variables. One approach to the problem is through Principal Component Analysis (PCA). The principal components achieved through transformation are ordered in decreasing order of importance. In various applications we can replace the set of p original variables by a smaller set of variables, provided that they contain a sufficient amount of information concerning the original observations. These new variables give us the possibility to reduce the dimension of the original data and sometimes reveal hidden structures. Furthermore, a transformation called whitening is often the first step in ICA. The whitening transformation is closely related to the PCA transformation.

When facing multivariate data with a significant amount of dimensions, the PCA transformation is sometimes used to reduce the dimensions of the variables before the ICA transformation. This is due to the fact that simultaneous diagonalization of two scatter matrix functionals requires clearly distinct eigenvalues. The numerical softwares used by us make approximations and sometimes the eigenvalues do not seem to be distinct. Also the algorithms used to conduct the ICA transformation are usually much heavier compared to the PCA algorithm.

In this chapter we will review the PCA and whitening transformations and give the definition of an invariant coordinate system. Furthermore, the dimension reduction using PCA will be discussed. The different effects of the whitening and PCA transformations are clearly visible in Figure 3.1. We give the definitions for complex variables in this Chapter. The real versions are simply achieved by replacing the conjugate transposes with regular transposes.

3.1 Invariant Coordinate System

We defined affine equivariant location and scatter functionals in Section 2.2. In the context of multivariate testing the desired property is called invariance. Invariance means that the results of the statistical test do not change whenever we perform scaling or location shifts to the original data. By requiring affine invariance, we ensure that the results are not affected by the choice of the coordinate system. Let x denote a p variate random vector with a cdf F_x and let $X = [x_1, \ldots, x_n]$, where $x_1, \ldots, x_n$ is a random sample from the distribution F_x.

Definition 3.1.1 (Affine Invariance)**.** *A functional $Q(F_x)$ is affine invariant if*

$$Q(F_{Ax+b}) = Q(F_x),$$

for all full rank $p \times p$ matrices A and for all p variate vectors b.

A statistic $Q(X)$ is affine invariant if

$$Q(AX + b1_n^T) = Q(X),$$

for all full rank $p \times p$ matrices A and for all p variate vectors b.

A definition of an invariant coordinate system functional is given below.

Definition 3.1.2 (Invariant Coordinate System Functional)**.** *An invariant coordinate system functional is a non-singular $p \times p$ matrix valued function $G(F_x)$ satisfying*

$$G\left(F_{Ax+b}\right) = G\left(F_x\right) A^{-1},$$

for all full rank $p \times p$ matrices A and all p variate vectors b.

Furthermore, the sample statistic of the invariant coordinate system functional satisfies the following,

$$G\left(AX + b1_n^T\right) = G(X)A^{-1}, \tag{3.1}$$

for all full rank $p \times p$ matrices A and all p variate vectors b. Note that invariant coordinate system functional can be used to pre-process data to obtain affine invariant or equivariant statistical procedures. For more details, see Ilmonen et al. (2012b). Note that the definitions in this Section also hold for complex valued statistics and functionals.

3.2 Whitening

As already discussed, the ICA problem is significantly simplified if the observed output vectors are first whitened. The origin of the term "white" comes from the fact that the power spectrum of white noise is constant over all frequencies, similarly like the visible light spectrum of white light contains all colors. Sometimes the term sphered is used instead of white. The whitening procedure is simply decorrelation followed by scaling. First we subtract the mean vector to move the location center to the origin, then we rotate the data to jointly uncorrelate the variables and finally rescale the variables to have unit variances.

Let x be a complex p variate random vector with a cumulative distribution function F_x and let $X = [x_1, ..., x_n]$, where $x_1, ..., x_n$ are observations from the distribution F_x. The whitening transformation can be formulated the following way.

Definition 3.2.1 (Whitening Transformation)**.**

$$y = S_1^{-1/2}(F_x)(x - \mathbb{E}(x)),$$

where $S_1^{-1/2}(F_x)$ is a square root of the inverse of the covariance matrix of F_x. After the transformation the following are true:

$$\mathbb{E}(F_y) = 0_p \quad \text{and} \quad S_1(F_y) = I_p.$$

All the directions have the same variance after the whitening transformation. The whitening transformation makes normally distributed variables independent, whereas for other distributions the transformation uncorrelates the variables. Note that $S_1^{-1/2}(F_x)$ is defined only up to an orthogonal transformation: if $Q\Lambda Q^T = I_p$, then also $(VQ)\,\Lambda\,(VQ)^* = I_p$ for all orthogonal matrices V. We can define the inverse square root of a matrix uniquely using the spectral decomposition.

Definition 3.2.2 (Spectral Decomposition)**.** *Let $S(F_x)$ be a $p \times p$ positive definite and symmetric scatter functional. Then $S(F_x)$ can be factorized as*

$$S(F_x) = Q\Lambda Q^*$$

where Λ is a diagonal matrix containing the eigenvalues of $S(F_x)$ in descending order and Q is an orthogonal matrix containing the corresponding eigenvectors as columns.

The spectral decomposition always exists for the covariance matrix since $S_1(F_x)$ is a positive semidefinite and conjugate symmetric matrix, see Horn

and Johnson (1985) for more details. If we further assume that all the components of x are non-degenerate and none of the components is fully linearly dependent of the other components, the covariance matrix is then positive definite. Now we can define the inverse square root as

$$S_1(F_x)^{-1/2} = Q\Lambda^{-1/2}Q^*, \tag{3.2}$$

where $S_1(F_x)^{-1/2}$ is uniquely defined. The following theorem holds for the invariance of the matrix inverse square root of a scatter functional.

Theorem 3.2.1. *If $S(F_x)$ is a scatter functional then*

$$S^{-1/2}\left(F_{Ax+b}\right) = OS^{-1/2}(F_x)A^{-1},$$

for some orthogonal $O = O\left(F_x, A\right)$.

The proof of Theorem 3.2.1 is pretty straightforward, see Appendix B.1. The real versions of the definitions are achieved by replacing the conjugate transpose with the regular transpose. For more details in the real case, see Ilmonen et al. (2012b).

Hereby, the coordinate system obtained by the whitening transformation is not affine invariant. It only holds that

$$S_1^{-1/2}\left(F_{Ax+b}\right)\left(Ax + b - \mathbb{E}\left(Ax + b\right)\right) = OS_1^{-1/2}\left(F_x\right)\left(x - \mathbb{E}\left(x\right)\right), \tag{3.3}$$

for some orthogonal matrix $O(F_x, A)$. For the sample version it holds that

$$S_1^{-1/2}(Y)\left(Y - \mu_Y 1_n^T\right) = OS_1^{-1/2}(X)\left(X - \mu_X 1_n^T\right), \tag{3.4}$$

where $Y = AX + b1_n^T$, μ_i is the corresponding sample mean and $O(X, A)$ is some orthogonal matrix.

3.3 Principal Component Analysis

Similar to the whitening transformation, the Principal Component Analysis (PCA) transformation also creates a coordinate system where the variables are uncorrelated. The uncorrelated variables are called principal components. The difference to whitening is that in classical PCA the variables are not scaled to have unit variances. The PCA transformation creates linear combinations of the original variables such that the variance is maximized under the constraint of being orthogonal to the previous variables. The PCA transformation is obtained using the spectral decomposition of the covariance matrix.

Let x be a complex p variate random vector with a cumulative distribution function F_x and let $X = [x_1, ..., x_n]$, where $x_1, ..., x_n$ are observations from the distribution F_x.

$$S_1(F_x) = Q\Lambda Q^*, \tag{3.5}$$

where Λ is a diagonal matrix containing eigenvalues of $S_1(F_x)$ in decreasing order and Q is orthogonal matrix containing the corresponding eigenvalues as columns. The classical approach to PCA is to use the covariance matrix as $S_1(\cdot)$. However, $S_1(\cdot)$ can be replaced with any scatter matrix functional to obtain e.g. a more robust transformation. The new coordinates are called principal components, when we use the regular covariance matrix.

Definition 3.3.1 (Principal Components)**.** *The principal components of a p variate random vector x are obtained through the transformation*

$$y = Q^* (x - \mathbb{E}(x)),$$

where the columns of Q contain the eigenvectors of the covariance matrix of x.

The sample principal components are then

$$Y = \hat{Q}^* (X - \mu_X), \tag{3.6}$$

where $\hat{Q}$ contains the eigenvectors of the sample covariance matrix of X and μ_X is the sample mean. Now $S_1(y) = I_p$ and $\mathbb{E}(y) = 0_p$. If we use a different scatter matrix functional, we can for example call y the principal components with respect to the chosen scatter matrix functional.

If the components of x have completely different scales, it is recommended to scale the components before the PCA transformation. By scaling the components $x_i \leftarrow x_i/\sigma_i$, where σ_i^2 is the variance of x_i, we give all the components same weight in the analysis. This transformation corresponds to performing the PCA transformation using the correlation matrix instead of the covariance matrix. In general PCA does not make any distributional assumptions except the existence of second moments. The difference between PCA and whitening transformations is demonstrated in Figure 3.1 where the two transformations are applied to a sample of 200 bivariate normal observations. The larger variance in the first principal direction is lost after the whitening transformation. The PCA transformation only centers and rotates the data here. Thus, the first principal direction has a larger variance compared to the second principal direction.

One of the main applications for PCA is dimension reduction. Given a data set on p dimensions, we can use PCA to find a linear subspace of dimension d lower than p. Such reduced subspace attempts to maintain as much of the variability of the data as possible. The variances of the principal components are directly given by the eigenvalues of the covariance matrix. The sum of k first eigenvalues divided by the sum of all eigenvalues represents the proportion of total variance explained by k first principal components. Hereby, it is straightforward to calculate the amount of variance that is retained after the dimension reduction. In optimal situations, a small number of principal components is sufficient and the dimension reduction allows us to visualize multivariate data in 2D or 3D graphics.

The principal components are not independent in general. Furthermore, traditional PCA is a very nonrobust method. However, we can replace the covariance matrix and the mean vector with some more robust estimators to achieve a robust version of PCA. It is obvious that PCA is not an invariant coordinate system functional in the sense of Definition 3.1.2. Simple scaling of the variables can result in completely different principal components and hereby completely different interpretations. Thus, when performing PCA we should always be careful that the results are not simply artifacts of the chosen coordinates system.

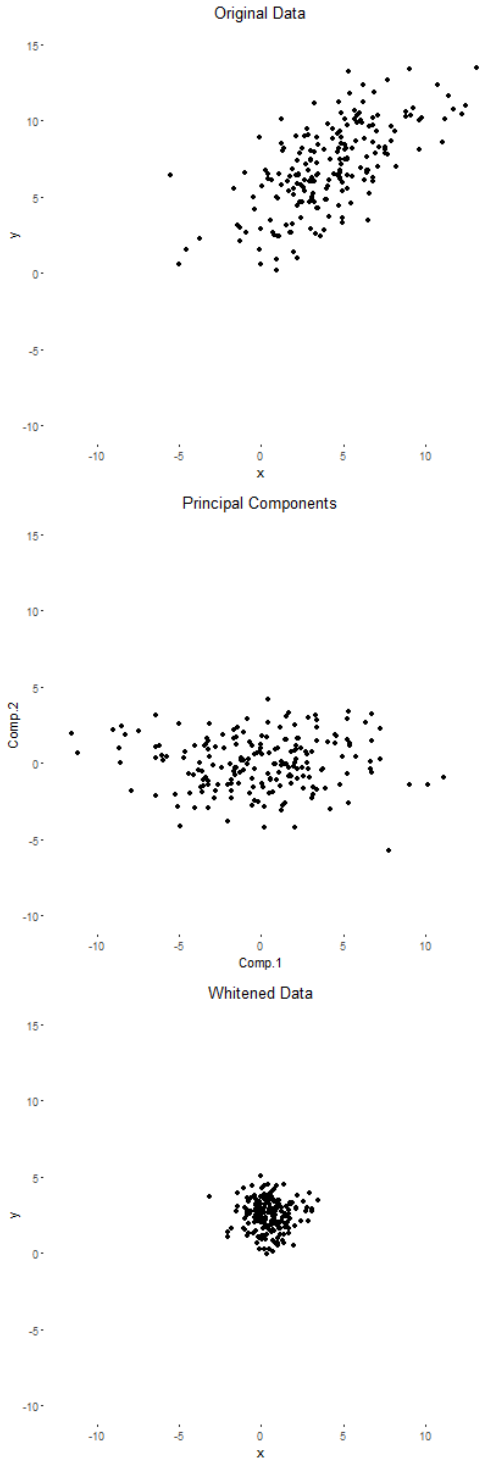


Figure 3.1: The top figure contains 200 observations simulated from a real bivariate normal distribution, the middle figure contains the principal components and the bottom figure contains the whitened coordinates of the original data.

Chapter 4

Affine Invariant Transformations

In this Chapter we introduce the independent component functionals based on the use of two scatter matrices. We achieve different transformations by choosing different $S_1(\cdot)$ and $S_2(\cdot)$ functionals. There are several other solutions to the independent component problem alongside FOBI and AMUSE, see Comon and Jutten (2010) and Hyvärinen et al. (2001).

FOBI and AMUSE are suited for different setting. The more optimal method depends on the situation. When we use FOBI, we assume i.i.d. data and hereby the order of the observations does not matter. Contrarily, when using AMUSE we assume that we have time series like data and construct our functionals based on the time structure of the data. Hereby, the complex FOBI has worse performance than AMUSE when considering complex time series data. Furthermore, we illustrate that AMUSE actually works for complex valued time series data. In this Chapter, we first give general theory that applies for both AMUSE and FOBI. Then we consider them individually. In the last Section of this Chapter, we generalize the Minimum Distance (MD) index for complex valued setting. This new formulation of MD index allows us to compare the complex FOBI and AMUSE.

4.1 ICS Functionals Based on the Use of Two Scatter Matrices

In this Section we discuss the simultaneous use of two scatter functionals. The theory behind the joint usage of two scatter functional has been developed quite recently. It was first discussed by Cardoso (1989). See also Oja et al. (2006) for the usage of two scatter matrices for real valued ICA and Ollila et al. (2008) for complex valued ICA.

A general approach to the problem without the IC-model assumption was

given by Tyler et al. (2009). The approach was named Invariant Coordinate Selection (ICS). The idea of ICS is to compare two scatter functionals $S_1(F_x)$ and $S_2(F_x)$ by solving $\Gamma(F_x)$ and $\Lambda(F_x)$ from the eigenvector-eigenvalue problem

$$S_1^{-1}(F_x)\, S_2(F_x)\, \Gamma(F_x)^* = \Gamma(F_x)\, \Lambda(F_x), \tag{4.1}$$

where F_x is the cdf of a complex valued p variate random vector x, Λ is a diagonal matrix containing the p eigenvalues of $S_1^{-1}(F_x)S_2(F_x)$ and the column vectors of $\Gamma(F_x)$ are the corresponding eigenvectors. Let $X = (x_1, \ldots, x_n)$ be a sample from the distribution F_x. Since the order, sign and length of the eigenvectors are not uniquely defined, we need to fix them. The eigenvectors are scaled using $S_1(\cdot)$ and ordered using $S_2(\cdot)$. We denote the values of the functionals from the empirical cdf as $\hat{S}_1 = S_1(F_n) = S_1(X)$ and $\hat{S}_2 = S_2(F_n) = S_2(X)$. Define $\hat{\Gamma} = \Gamma(X)$ and $\hat{\Lambda} = \Lambda(X)$ as the solution to the equations

$$\hat{\Gamma}\hat{S}_1\hat{\Gamma}^* = I_p \quad \text{and} \quad \hat{\Gamma}\hat{S}_2\hat{\Gamma}^* = \hat{\Lambda}, \tag{4.2}$$

where $\hat{\Lambda}$ is a diagonal matrix with diagonal elements $\hat{\lambda}_1 \geq \ldots > 0$. Note that the solution always exists. If $\lambda_1 > \ldots > \lambda_p > 0$, then the unmixing matrix Γ is uniquely defined up to multiplication with $L = \mathrm{diag}\left(\exp\left(\theta_1 i, \ldots, \theta_p i\right)\right)$. In the real case, replace L with a sign change matrix. In the case of non-distinct eigenvalues of $S_1^{-1}(F_x)S_2(F_x)$, see Tyler et al. (2009).

In this case, x is first uncorrelated and scaled by performing the whitening transformation, $y = S_1^{-1/2}(F_x)(x)$. Then PCA transformation is performed on y using $S_2(F_y)$. Actually, ICS gets its name from transformation that yields an invariant coordinate system in the sense that

$$\Gamma(F_{Ax+b}) = L(F_{Ax+b})\Gamma(F_{Ax+b})A^{-1}, \tag{4.3}$$

where $L = \mathrm{diag}\left(\exp(\theta_1 i), \ldots, \exp(\theta_p i)\right)$ fixes the the diagonal elements of $\Gamma(F_{Ax+b})$ to be positive real numbers. The new coordinates are independent with respect to two different measures of dependence. Since the modulus of each of the diagonal elements of L is equal to one, we can replace L with a sing change matrix if we assume real valued variables. If we want to obtain full affine invariance in the sense of Definition 3.1.2, we can standardize the eigenvectors of $S_1^{-1}(F_x)\, S_2(F_x)$ using two affine equivariant location functionals $T_1(\cdot)$ and $T_2(\cdot)$. However, we choose to fix the model by assuming that the diagonal elements of $\Gamma(F_{Ax+b})$ are positive real numbers. This is due to the asymptotic results in Ilmonen (2013). Fixing the argument of the independent components using the location functional complicates the deriving of asymptotic results. The matrix Γ can now be used in transforming the

random vector x to invariant coordinates under the IC-model assumption. The transformation solves the ICA problem.

4.2 FOBI

The Fourth Order Blind Identification (FOBI) functional is probably the first solution for the independent component problem, first introduced in Cardoso (1989). We obtain the FOBI functional $\Gamma(F_x)$ if the scatter functionals $S_1(F_x)$ and $S_2(F_x)$ in Formula 4.1 are chosen to be the covariance matrix and the scatter matrix based on fourth moments respectively. FOBI is computationally one of the most efficient ICA methods. Note that, FOBI requires distinct kurtosis values for the independent components. If some of the independent component have identical kurtosis values, we can still estimate those that have distinct kurtosis values.

Let X be a random sample from the complex IC model. We can construct the complex valued FOBI estimate $\hat{\Gamma}(X)$ using the following steps:

1. Whitening transformation: $Y_i \leftarrow S_1(X)^{-1/2}(X_i - \mu_X)$, where $A^{1/2}$ is the conjugate symmetric square root of A and μ_X is the sample mean.
2. Compute the scatter matrix based on fourth moments: $S_2(Y)$
3. Finding eigenvalue - eigenvector decomposition of $S_2(Y) = UDU^*$, where D is a diagonal matrix and U is orthogonal.
4. Compute the final unmixed observations: $\hat{Z} = U^*Y$.

Note that for a real A, $A^* = A^T$. The asymptotic properties of FOBI have been studied in the real case in Ilmonen et al. (2010a) and in the complex case in Ilmonen (2013). Under the IC-model assumption, consistency of the FOBI estimator is obtained if the fourth moments of the independent components exists as finite quantities and are distinct. Moreover, $\sqrt{n}$-consistency and asymptotic normality follow it the eight moments of the independent components are finite.

4.3 AMUSE

The Algorithm for Multiple Unknown Signals Extraction (AMUSE) was first proposed by Tong et al. (1990). We obtain the AMUSE functional $\Gamma(F_x)$ if the scatter functionals $S_1(F_x)$ and $S_2(F_x)$ are the covariance matrix and

the autocovariance matrix with a chosen time lag τ. Usually, AMUSE performs well with the choice of τ being 1. In order to obtain positive definite scatter functionals and estimates, we only apply symmetrized autocovariance matrices.

Let $X = [x_1, \ldots, x_n]$ where $x_1, \ldots, x_n$ is a time series data from the distribution F_{x_t}. We can construct the complex valued AMUSE estimate $\hat{\Gamma}(X)$ using the following steps:

1. Whitening transformation: $Y_{t_i} \leftarrow S_1(X)^{-1/2}(X_{t_i} - \mu_{X_t})$, where $A^{1/2}$ is the conjugate symmetric square root of A and μ_{X_t} is the sample mean.

2. Compute the autocovariance matrix: $S_\tau(Y_t)$, for a chosen τ.

3. Symmetrizing the autocovariance matrix:
   $S_\tau(Y_t)_{\text{symm}} = (S_\tau(Y_t) + S_\tau(Y_t)^*)/2$.

4. Finding eigenvalue - eigenvector decomposition of $S_\tau(Y_t)_{\text{symm}} = UDU^*$, where D is a diagonal matrix and U is orthogonal.

5. Compute the final unmixed time series: $\hat{Z}_t = U^* Y_t$.

The asymptotic properties of AMUSE have been studied in the real case in Miettinen et al. (2012). Under the IC-model assumption, consistency of the AMUSE estimator is obtained if the second moments of the independent components exists as finite quantities and are distinct. Moreover, $\sqrt{n}$ consistency and asymptotic normality follow if the fourth moments of the independent components are finite.

The asymptotic behavior of the complex valued AMUSE functional has not yet been considered in the literature. However, the asymptotic results in the complex case will be considered in Lietzén et al. (2016). The simulation results in Chapter 5 already hint that the AMUSE estimator is asymptotically normal if the fourth moments of the independent components are finite.

The behavior of complex valued AMUSE, FOBI and PCA are illustrated in Figures 4.1 and 4.2. In Figure 4.1 we have a small simulated example that illustrates the separation of bivariate complex valued distributions. The first row contains the original simulated data which is then mixed with a complex valued mixing matrix. The mixed distributions are on the second row. Then, we use PCA, FOBI and AMUSE to separate the signals. We see that only AMUSE manages to find the original distributions. This is due to the spatial dependence between the observations in the Φ distribution. The Φ distribution is rotated compared to the original distribution since the functional $\Gamma(\cdot)$ is unique up to permutations and phases.

In Figure 4.2 we apply AMUSE, FOBI and PCA to a bivariate complex normal distribution that is on the first row. We see that FOBI and AMUSE transformation lose the information regarding the variation on the first component. This is due to normally distributed variables being independent after performing the whitening transformation. However, the PC transformation retains the variation in the first component.

4.4 Minimum Distance Index

Let $X = (x_1, \ldots, x_n)$ be a random sample from the independent component model M1 or M2 defined in Section 2.6.1 with some choice of Ω and z. An estimate of $\Gamma(F_x)$ is obtained, when we apply the functional to the sample cumulative distribution function F_n. We denote the estimate by $\hat{\Gamma}$, $\Gamma(F_n)$ or $\Gamma(X)$. The gain matrix $\hat{G} = \hat{\Gamma}\Omega$ is used to compare the performance of different estimators. We have multiple ways of comparing matrices $\hat{G}$ that converge to a population value that is dependent on the functional Γ and the choice of Ω and z. We can for example use canonical parametrization, adjusted functionals or the Minimum Distance (MD) index, see Ilmonen et al. (2012a).

In this Thesis we measure the performance of the estimators using MD index. Furthermore, we consider the MD index in the complex valued case that to our knowledge has not been considered before. Let $\mathcal{C} = \{DP : \ P \text{ is a } p \times p \text{ permutation matrix, } D \text{ is a } p \times p \text{ diagonal matrix}\}$. The sets $\{CA : C \in \mathcal{C}\}$ partition the set of $p \times p$ matrices into equivalence classes. If $B \in \{CA : C \in \mathcal{C}\}$, notation $A \sim B$ is used. The shortest squared distance between the set $\{CA : C \in \mathcal{C}\}$ of matrices that are equivalent to A and I_p is given by

$$D^2(A) = \frac{1}{p-1} \inf_{C \in \mathcal{C}} ||CA - I_p||_F^2 , \tag{4.4}$$

where $|| \cdot ||_F$ is the Frobenius norm.

Remark 4.4.1. *Note that $D^2(A) = D^2(CA)$ for all $C \in \mathcal{C}$.*

Theorem 4.4.1. *Let A be any $p \times p$ full rank matrix. The shortest squared distance $D^2(A)$ fulfills the following four conditions given below:*

1. $1 \geq D^2(A) \geq 0$,
2. $D^2(A) = 0$ *iff* $A \sim I_p$,
3. $D^2(A) = 1$ *iff* $A \sim 1_p a^T$ *for some p variate vector a, and*

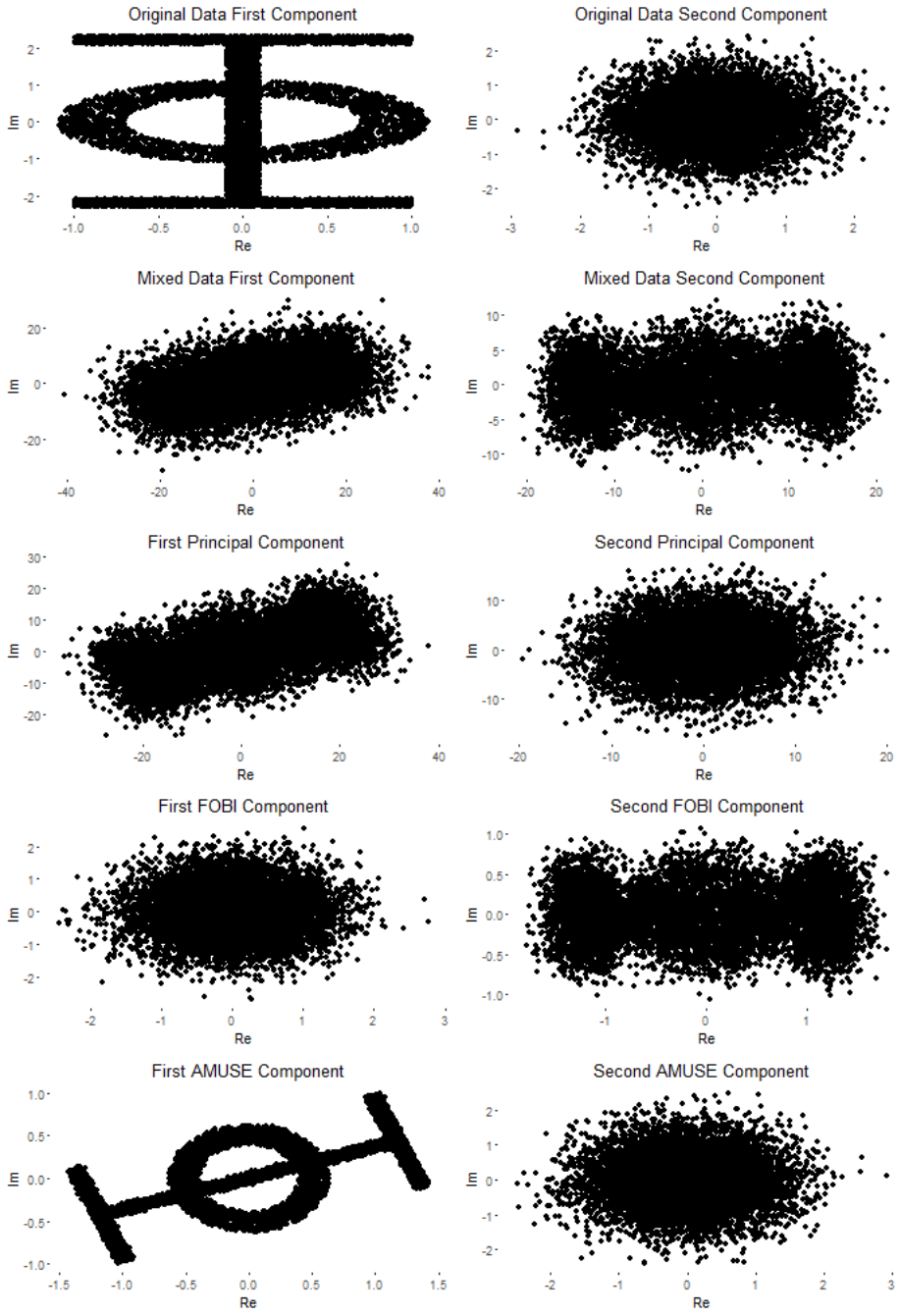


Figure 4.1: The top row contains the original observations. The second row contains the original observations mixed with a complex mixing matrix. The last three rows contain the estimated principal and independent components obtained from PCA, FOBI and AMUSE.

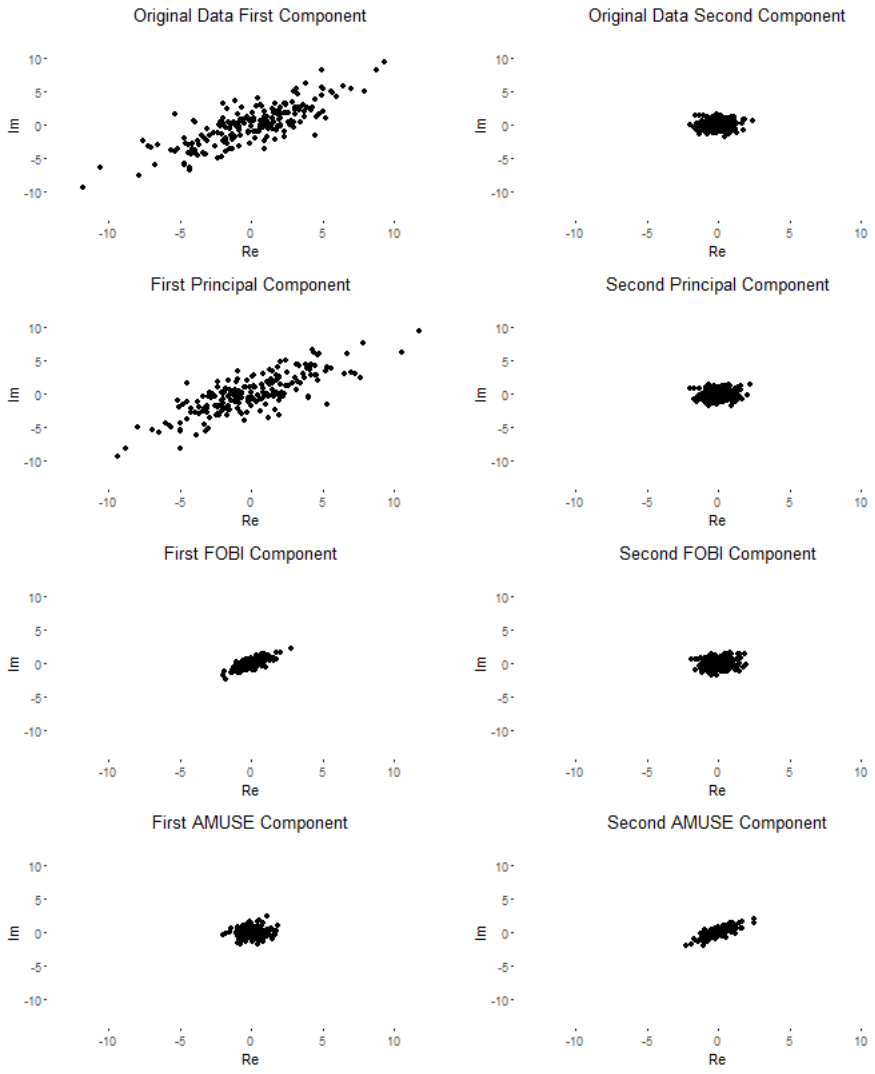


Figure 4.2: The first row contains the two dimensions of a sample from a bivariate complex normal distribution. The rest of the rows contain data after PCA, FOBI and AMUSE transformations.

4. *the function* $c \to D^2\left(I_p + c \text{ off}\left(A\right)\right)$ *is increasing in* $c \in [0,1]$ *for all matrices* A *such that* $A_{ij}^2 \leq 1$, $i \neq j$.

Let $X = [x_1, \ldots, x_n]$ be a random sample from a distribution F_x, where x follows the model defined in 2.6.1 with an unknown mixing matrix Ω. Let $\Gamma(F)$ be a separation functional. Hereby, $D^2\left(\Gamma\left(F_x\right)\Omega\right) = 0$. The unmixing matrix estimate based on the functional $\Gamma\left(F_x\right)$ is

$$\hat{\Gamma} = \hat{\Gamma}\left(X\right) = \Gamma\left(F_n\right), \tag{4.5}$$

where F_n is the empirical cumulative distribution function based on X. The shortest distance between the identity matrix and the set of matrices $\left\{C\hat{G} : C \in \mathcal{C}\right\}$ equivalent to the gain matrix $\hat{G} = \hat{G}\Omega$ is given in the following definition.

Definition 4.4.1 (Minimum Distance Index)**.** *The minimum distance index for* $\hat{\Gamma}$ *is*

$$\hat{D} = D\left(\hat{\Gamma}\Omega\right) = \frac{1}{\sqrt{p-1}} \inf_{C \in \mathcal{C}} \left|\left|C\hat{\Gamma}\Omega - I_p\right|\right|_F.$$

It follows from Theorem 4.4.1 that $1 \geq \hat{D} \geq 0$, and $\hat{D} = 0$ only if $\hat{\Gamma} \sim \Omega^{-1}$. Furthermore, $\hat{D} = 1$ is obtained when all the row vectors of $\hat{\Gamma}\Omega$ have the same direction. Hereby the value of the minimum distance index is easy to interpret, values near 0 are associated with successful separation. Note that $D(\hat{\Gamma}\Omega) = D(C\hat{\Gamma}\Omega)$ for all $C \in \mathcal{C}$. Additionally,

$$D\left(\Gamma\left(XA^T\right)A\Omega\right) = D\left(\Gamma(X)\Omega\right), \tag{4.6}$$

due to affine equivariance of Γ. The minimum distance index is compared to other performance estimators in Nordhausen et al. (2011c).

The minimization over all choices $C \in \mathcal{C}$ is done by two steps.

Lemma 4.4.1. *Let* $\mathcal{P}$ *denote the set of all* $p \times p$ *permutation matrices. Let* $\hat{G} = \hat{\Gamma}\Omega$ *and let* $\tilde{G}_{ij} = |\hat{G}|_{ij}^2 / \sum_{k=1}^p |\hat{G}|_{ik}^2$. *Now the minimum distance index can be written as*

$$\hat{D} = D(\hat{G}) = \frac{1}{p-1}\left(p - \max_{P \in \mathcal{P}}\left(\text{tr}\left(P\tilde{G}\right)\right)\right)^{1/2}.$$

The maximization problem

$$\max_P \left(\text{tr}(P\tilde{G})\right) \tag{4.7}$$

over all permutation matrices P can be seen as a linear sum assignment problem (LSPA). We solve the LSPA using the Hungarian method, see Papadimitriou and Steiglitz (1982). The proof of Lemma 4.4.1 for complex valued numbers is in Appendix B.2. Note that due to the absolute values, $\tilde{G}$ is guaranteed to be a real number. Hereby, we can calculate the minimum distance for both real and complex numbers using the formulation in Lemma 4.4.1. The R-implementation for the minimum distance index that can handle complex numbers is in Appendix C.3. The code is a modified version of the minimum distance index function in the R-package JADE by Nordhausen et al. (2011a).

Chapter 5

Simulation Study

To demonstrate the importance of choosing the right estimator for the right setting, we conduct a small simulation study. Furthermore, we wanted to asses the effect of choosing the parameter τ. We compare the estimates of complex AMUSE with complex FOBI in two different settings using the Minimum Distance (MD) index. The first setting contains observations from a time series and the second from a non-time series data. We illustrate the difference in performance in these two different settings (and with different parameters τ). The performance of real valued AMUSE in different settings is considered by Miettinen et al. (2012) and the performance of real FOBI is considered in Ilmonen et al. (2010b).

We calculated the MD index 5000 times for three variate mixed signals. The signals had 10^5 observations. The first setting contained three independent stationary AR(2) processes with normally distributed innovations. We made the time series complex valued such that the components of the real part was simulated from the independent AR(2) processes. The imaginary part was then formed as a linear combination of the corresponding real component with an added normally distributed innovation. The real and complex parts were made dependent on each other to illustrate the value of complex valued unmixing compared to the real case. If there is no dependency between the complex and real part in the p dimensional case, it is more convenient to perform the unmixing using $2p$ dimensional real functionals.

The second setting contained observations from beta distribution, uniform distribution and Student's t-distribution with degree of freedom 10 and all of the components had normal innovations. Here the complex and real part of the components were simulated independently from the corresponding distribution. This setting was made to illustrate the better performance of FOBI in the i.i.d. case.

In both settings, the sources were mixed with an identity matrix. Having

a trivial mixing matrix does not affect the performance of the estimator since we are using affine equivariant methods. The results for the first setting in Figure 5.1 show that for a time series data the AMUSE has clearly better MD index values compared to FOBI. The effect of the choice of the parameter τ is also illustrated here. AMUSE performs better with small values of τ and the MD index distribution starts to shift to 1 as τ grows. FOBI does not perform well here since the components are not i.i.d. observations. It is clear that AMUSE is the better choice here in the complex valued time series setting.

In the i.i.d. setting in Figure 5.2 we see that FOBI performs better. Now the observations are i.i.d. resulting in the good performance of FOBI. Furthermore, since there is no autocovariance structure, the choice of the parameter τ does not have an effect on the performance of AMUSE. Hereby, in this i.i.d. setting FOBI is the better choice.

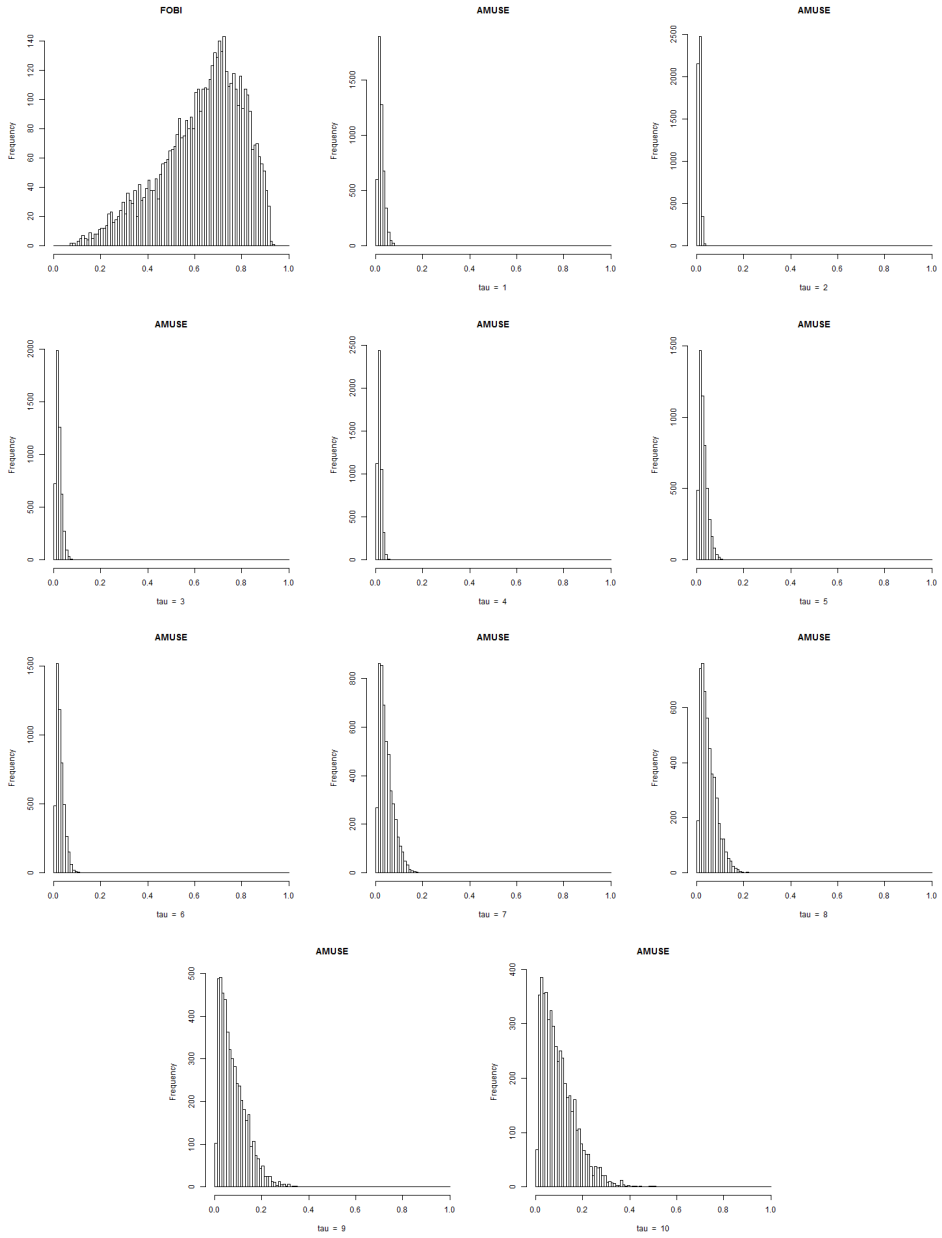


Figure 5.1: MD index calculated for three independent complex valued AR(2) processes with normally distributed innovations.

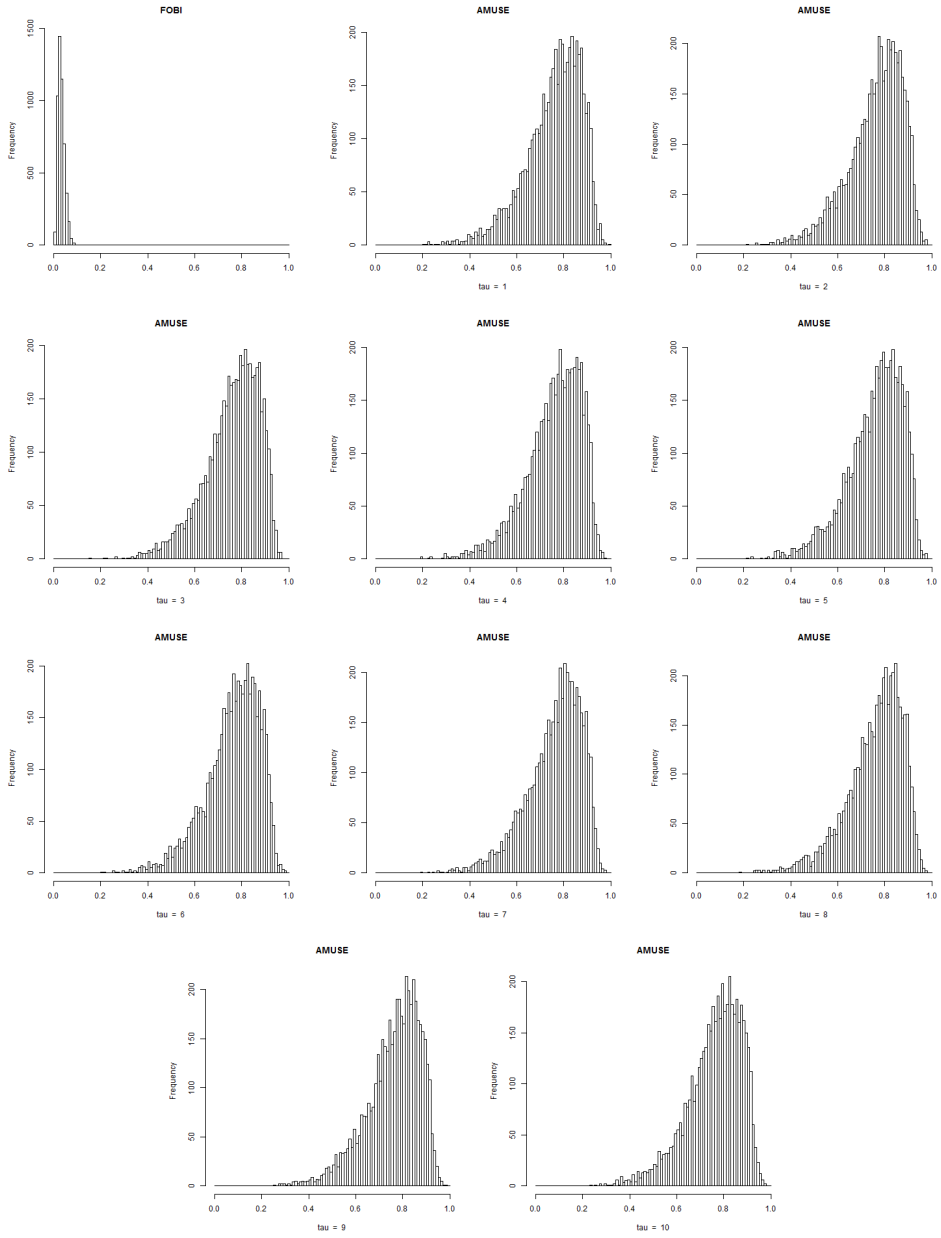


Figure 5.2: MD index calculated for complex valued i.i.d. observations where the first component is t-distributed, the second is beta distributed and the third is uniformly distributed. All the components have normally distributed innovations.

Chapter 6

Data Example

There are countless examples of real valued ICA in literature. Furthermore, ICA has been mostly considered in the complex valued case involving a transformation to the frequency domain. However, in this thesis we will be considering two example that involve naturally complex valued signals.

6.1 Image Source Separation

Separation of mixed and overlapping images is a common problem in image processing. A real life example is the separation of overlapped fingerprints obtained from a crime scene, where we observe a mixture of two or more fingerprints, see Singh et al. (2006).

We showcase the separation properties of complex AMUSE and FOBI by generating three images, seen in Figure 6.2a. We then mix the images with a random complex valued mixing matrix, the mixed images are in Figure 6.2b. After that we try to separate the original images from the mixtures using complex AMUSE and FOBI. The separated images are in Figures 6.2d and 6.2c. The R code for running the whole data example is in Appendix C.2. To our knowledge, complex valued AMUSE or FOBI has not been considered in such setting before.

The images were made complex valued by transforming the RGB values of each pixel to the complex plane using known topological transformations. However, we only used the colors on the surface of the RGB cube that is illustrated in Figure 6.1. By considering only the colors on the surface of the cube, we could define bijective transformations between the complex plane and the cube surface.

Hereby, our original data D is a $h \times w \times c$ multidimensional real valued array (tensor) containing linear mixtures of some figures, where h is the

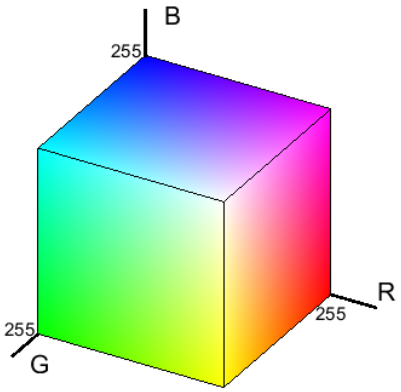


Figure 6.1: The RGB color cube generated using MATLAB.

height and w is the width of the original figure. Here the unit of measure is number of pixels. Every (h, w) coordinate has a corresponding c value that contains the R, B and G values for that pixel. Note that min(R,B,G) = 0 and max(R,B,G) = 255 and for the purposes of our transformations we scale the RBG values to be between 0 and 1. The first step of our algorithm is to ensure that all the pixels are colors that are on the surface of the RBG cube. We projected every point from the inside to the surface of the cube the following way:

1. $\mu_{max} = \max(\text{RBG}), \quad \mu_{min} = \min(\text{RBG})$
2. $\Delta = 255 - \mu_{max}$

3a. IF($\Delta \leq \mu_{min}$): $\quad \text{RBG}(\mu_{max}) \leftarrow 255$

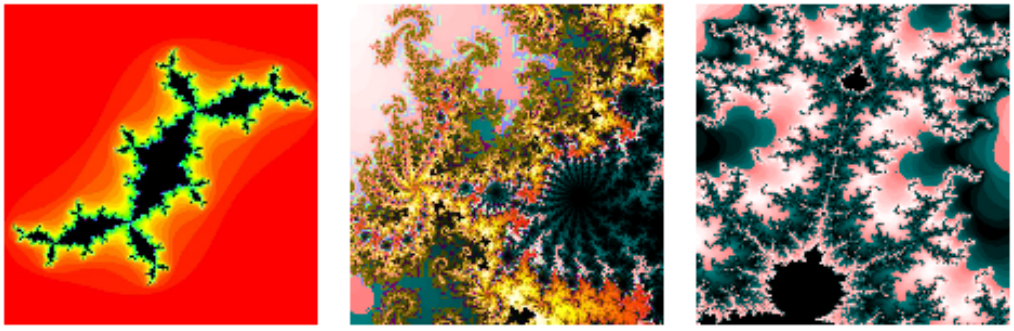
3b. ELSE: $\quad \text{RBG}(\mu_{min}) \leftarrow 0,$

where RBG = (R,B,G) $\in \mathbb{R}^3$. This preliminary step ensures that we only have to deal with colors on the surface of the cube. We then transform the cube surface into a sphere. After that, we can use the known stereographic projection that projects a sphere onto a plane. The steps are done in reverse order when we transform from the complex plane onto the RBG cube surface. The formulas for the transformations and inverse transformations are in Appendix A.3, A.4 and the corresponding R codes are in Appendix C.1. We can use the R function to generate a data example for any set of three figures in .jpeg format. However, high quality photographs do not generally look good after all the colors are transformed to the surface of the RBG cube.

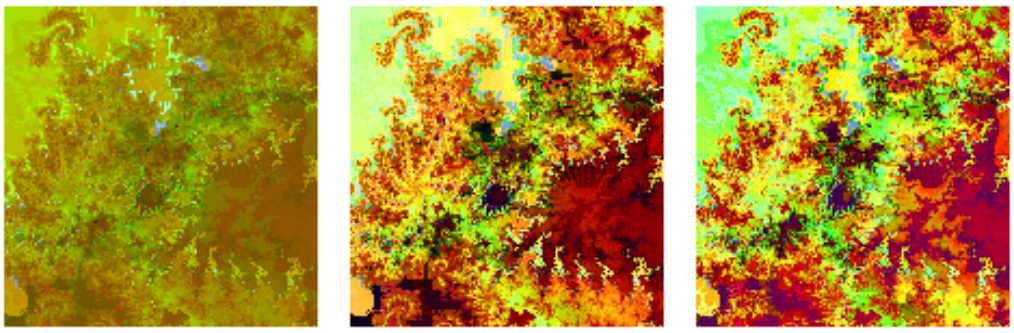
After the transformations, we have a $h \times w$ complex valued matrix for every image. We then vectorize the matrices and combine the columns resulting in a $n \times p$ complex valued matrix, where p is the number of images

and $n = h \cdot w$. We have considered an alternative vectorization, denoted by vec_s. In this alternative vectorization we vectorize matrices such that we start from the first elements in odd rows and from the last elements in even rows. Now the spatial dependencies of the image are highlighted more in the vectorized matrix, since pixels that are close to each other in the original image are also close to each other in the vectorized matrix. The purpose of the alternative vectorization is to provide better estimates for the autocovariance matrix in the AMUSE algorithm. The choice of vectorization has no effect on FOBI.

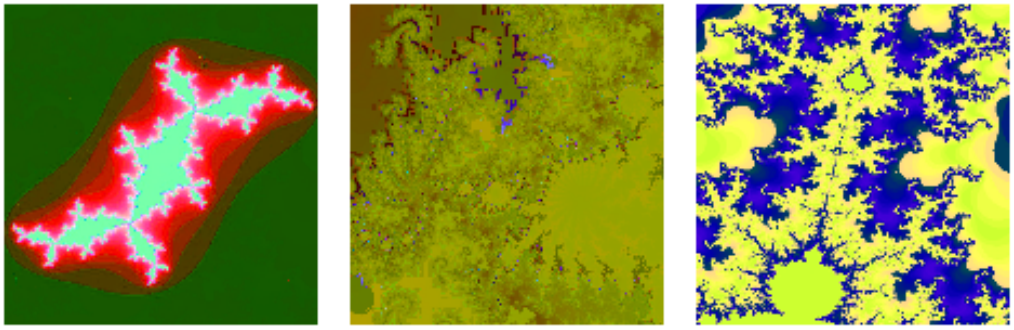
After the vectorization, we can perform the AMUSE and FOBI transformations for the mixed data. Both of the methods seem to work relatively well here. When comparing Figures 6.2a, 6.2c and 6.2d, we see that the shapes are retained but the colors seem to be different when comparing to the original images. This is due to the fact that the transformations are uniquely defined up to scale, sign and phase as we discussed in Section 4.1. The choice of vectorization does not seem to affect the AMUSE transformation in this example. We can see in Figure 6.4 that we get almost identical results with both vectorizations. Furthermore, the effect of τ is illustrated in Figure 6.3. We see that the AMUSE estimates start to get worse as τ grows. Furthermore, it is surprising that AMUSE and FOBI work in this example. The vectorized data does not fulfill the stationarity conditions.



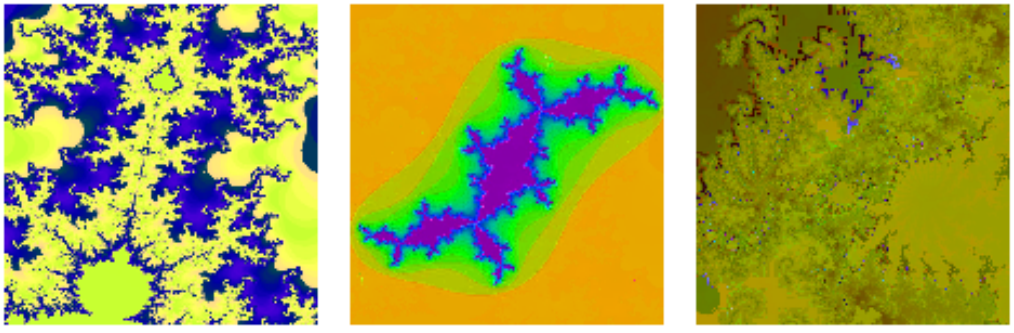
(a) Original images.



(b) Mixed images.

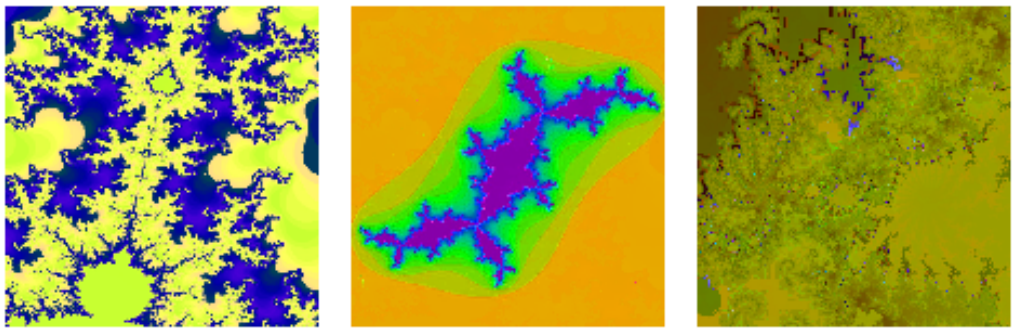


(c) Unmixed images using complex FOBI.

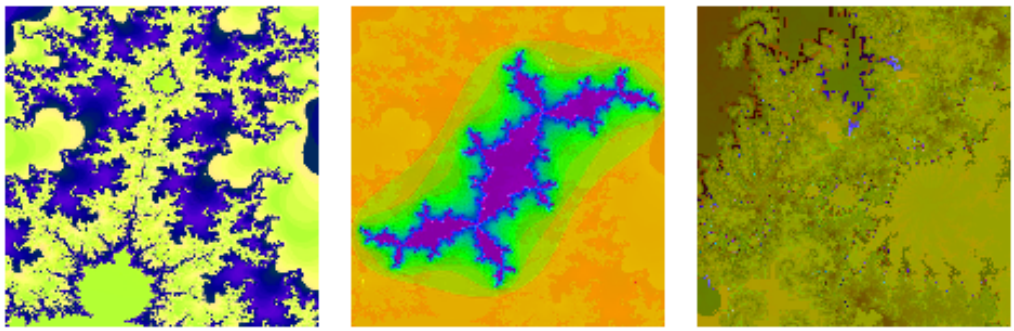


(d) Unmixed images using complex AMUSE with τ=1 and vec_s.

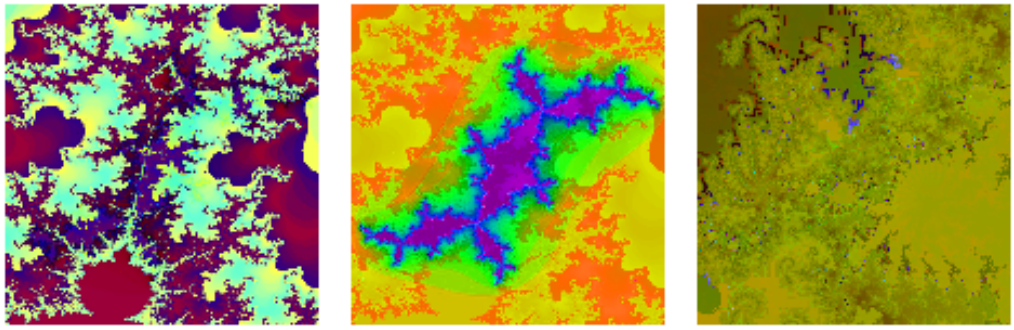
Figure 6.2: Estimates using FOBI and AMUSE.



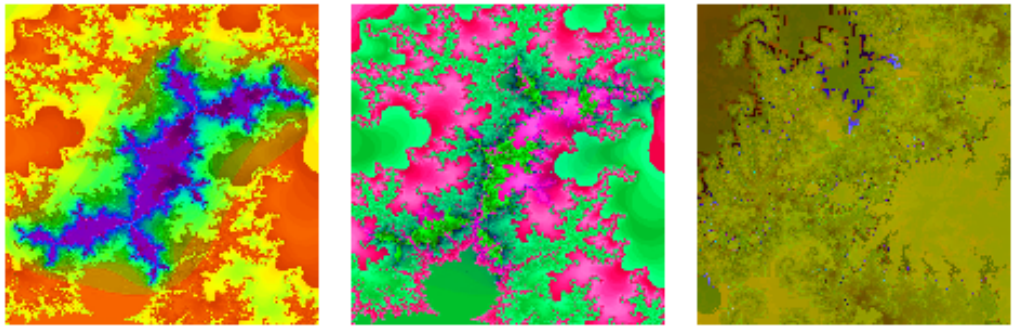
(a) τ=1



(b) τ=5

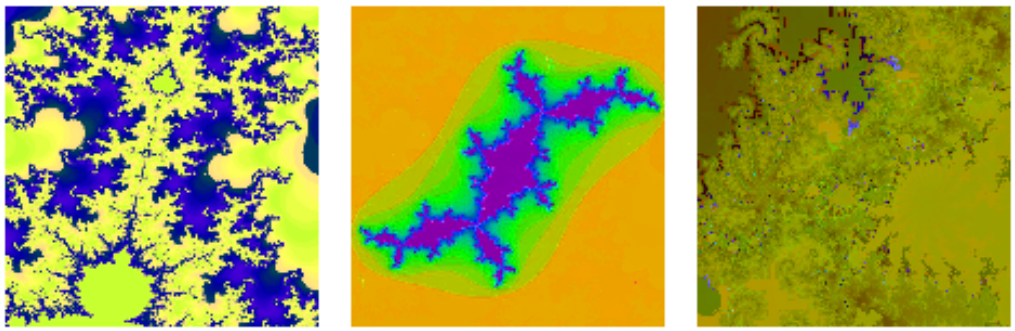


(c) τ=10

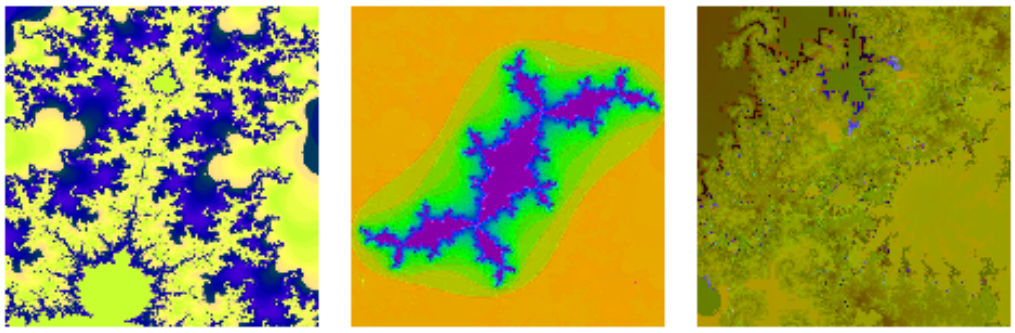


(d) τ=25

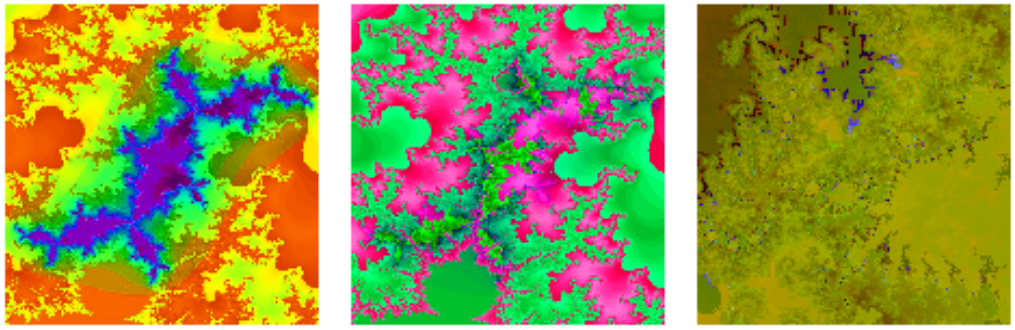
Figure 6.3: AMUSE estimates with different choices of τ using vec_s.



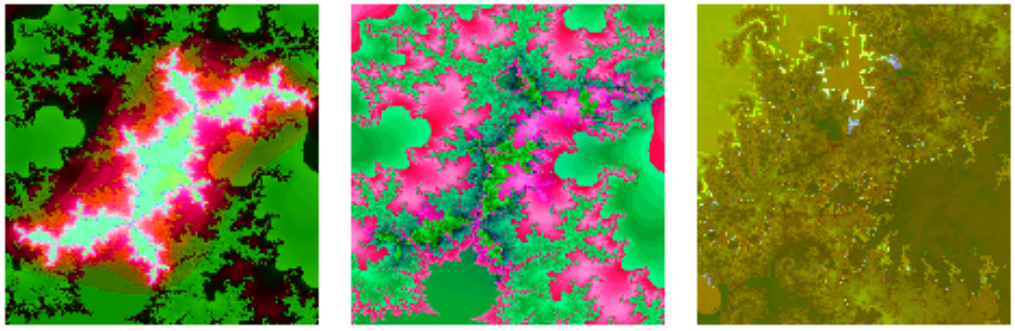
(a) τ=1 and vec_s



(b) τ=1 and regular vectorization



(c) τ=25 and vec_s



(d) τ=25 and regular vectorization

Figure 6.4: Comparison of AMUSE estimates with different vectorizations.

6.2 Functional Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) is a prominent non-invasive neuroimaging technique that is commonly used in clinical routine and advanced brain research. In functional Magnetic Resonance Imaging (fMRI) we measure changes in brain activity. In an activated area of the brain, the consumption of oxygen increases. At the same time, the blood flow to the activated areas increases. The magnetic properties of these activated areas differ from the rest of the brain and hereby we can measure them using a MRI scanner. The data produced is noisy, highly variating between subjects, massive in size and highly correlated both spatially and temporally, see Lange (2003). Statistical analysis is in a key role to make the datasets interpretable. The data sets resulting from an fMRI experiment are usually enormous. We have a time series of a $p \times p \times p$ mm^3 resolution images where the accuracy depends on the experiment. Usually one of these images contains more than 10^5 elements. Furthermore to capture the brain activity, these time between two images is set to be very small and the experiments are usually long compared to the collection rate.

Various preprocessing steps are required to correct the functional images. The images are corrected to eliminate the effects of possible head movement and for group studies the individual images are transformed to a common referential. The measuring process has many parts and is very technical. We end up with complex valued time series data due to the use of Fourier and inverse Fourier transformations. It is stated in Adali and Calhoun (2007) that most fMRI analysis techniques discard the phase of the data and consider only the magnitudes. The paper above has examples, where some brain activity is only detected when also the phase of the data is utilized.

Several methods have been proposed for the statistical analysis of the preprocessed sets of functional data. The most common approach is to use regression techniques, see Friston et al. (1994) and Bullmore et al. (1996). A common alternative approach to the problem is based on independent component analysis. The goal is to find the independent components that are related to a specific activity, for example moving an arm or a leg. At the same time, we separate the uninteresting noise components. The amount of applications would be enormous, if we could separate the parts of the brain that activate when the test subject wants to move a specific limb. This would allow for example a completely new type of robotic limbs or remote control of robots. The main problem of ICA is related to computational power and accuracy, which is mainly due to the enormous data size. An interesting approach to the problem could be a tensor valued complex ICA.

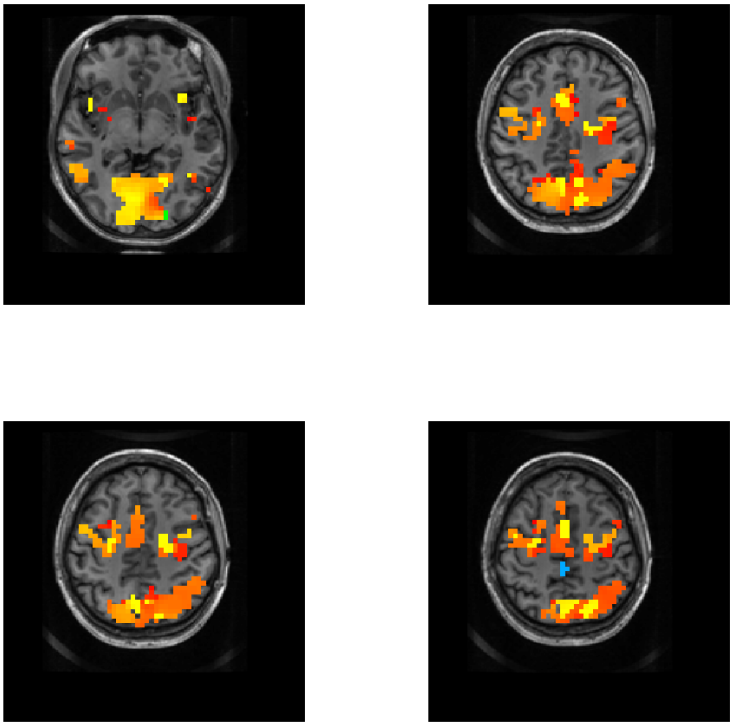


Figure 6.5: Functional MRI images. The colors indicate activity in the specific part of the brain.

Since suitable real data examples related to medical data are difficult to find, we decided to use a simulated example. Unfortunately, we did not find a dataset with both the phase and magnitude of the data. Note that a package exists in R for generating simulated fMRI data. The package neuRosim (Welvaert et al. (2011)) allows us to generate fMRI data with a large variety of activation models and noise structures. Unfortunately, the package does not yet allow the generation of complex valued fMRI data consisting of both magnitude and phase data. It is listed as future work in Welvaert et al. (2011). We are planning to test our method for the complex valued data when the package gets updated or when we find a suitable real dataset.

We have a simulated example in Figure 6.5. The example is generated using R packages AnalyzeFMRI and fmri. The image shows a potential independent component that we could find. The colors in the figure represent brain activity. The parts that do not have a color have activity lower than a threshold that we set.

Chapter 7

Further Research

The work of this Thesis will be continued in Lietzén et al. (2016). We will provide asymptotic results for the AMUSE transformation in that paper. Furthermore, in that paper the complex version of the minimum distance index will be discussed with more examples. An interesting research question would be to formulate the complex AMUSE transformation for settings that require less strict stationarity assumptions. Furthermore, the formulation and asymptotic properties of complex tensor valued ICA provides grounds for future research.

Appendix A

Formulas

A.1 Moments

A.1.1 Generalized Gamma Distribution

The generalized Gamma distribution contains many well known distributions. Let x follow the generalized Gamma distribution with a probability density function f_x,

$$f_x = \frac{ac\,(ax)^{bc-1}\exp\left(-(ax)^c\right)}{\Gamma(b)},$$

where a, b and c are real positive parameters. Furthermore, x can only have values larger or equal to zero. The relation of the generalized Gamma distribution to other distribution is in Table A.1. The first moment (expected

Table A.1: Generalized Gamma distribution with fixed parameters.

Distribution	a	b	c
Generalized Gamma	a	b	c
Gamma	a	b	1
χ^2	$1/2$	$n/2$	1
Exponential	$1/\alpha$	1	1
Folded Standard Normal	$1/\sqrt{2}$	$1/2$	2
Maxwell	$1/\left(\alpha\sqrt{2}\right)$	$3/2$	2
Rayleigh	$1/\left(\alpha\sqrt{2}\right)$	1	2
Weibull	$1/\alpha$	1	η

value) is

$$\begin{aligned}
&\int_0^\infty x f_x dx = \int_0^\infty x \frac{ac\,(ax)^{bc-1}\exp\left(-(ax)^c\right)}{\Gamma(b)} dx \\
&= \frac{c}{\Gamma(b)} \int_0^\infty (ax)^{bc} \exp\left(-(ax)^c\right) dx \qquad (u = a^c) \\
&= \frac{cu^b}{\Gamma(b)} \int_0^\infty x^{bc} \exp\left(-ux^c\right) dx = \frac{cu^b}{\Gamma(b)} \frac{1}{c} u^{-\frac{bc+1}{c}} \Gamma\left(\frac{bc+1}{c}\right) \\
&= \frac{\Gamma\left(b+\frac{1}{c}\right)}{a\Gamma\left(b\right)} = \mu.
\end{aligned}$$

Since the generalized Gamma distribution is not symmetric, the odd central moments can differ from zero. The second central moment (variance) is

$$\begin{aligned}
&\int_0^\infty (x-\mu)^2 f_x dx = \int_0^\infty (x-\mu)^2 \frac{ac\,(ax)^{bc-1}\exp\left(-(ax)^c\right)}{\Gamma(b)} dx \\
&= \frac{ca^{bc}}{\Gamma(b)} \int_0^\infty (x-\mu)^2 x^{bc-1} \exp\left(-(ax)^c\right) dx \qquad \left(C = \frac{ca^{bc}}{\Gamma(b)}\right) \\
&= C \int_0^\infty \left(x^{bc+1} - 2x^{bc}\mu + x^{bc-1}\mu^2\right) \exp\left(-(ax)^c\right) dx \qquad (u = a^c) \\
&= C \int_0^\infty x^{bc+1} \exp\left(-ux^c\right) dx - 2\mu \frac{cu^b}{\Gamma(b)} \int_0^\infty x^{bc} \exp\left(-ux^c\right) dx \\
&+ C\mu^2 \int_0^\infty x^{bc-1} \exp\left(-ux^c\right) dx \\
&= C \frac{u^{-b-\frac{2}{c}} + \Gamma\left(b+\frac{2}{c}\right)}{c} - 2\mu^2 + C\mu^2 \frac{\Gamma(b)}{u^b c} \\
&= \frac{ca^{bc}}{\Gamma(b)} \frac{a^{-(bc+2)} + \Gamma\left(b+\frac{2}{c}\right)}{c} - 2\left(\frac{\Gamma\left(b+\frac{1}{c}\right)}{a\Gamma\left(b\right)}\right)^2 + \frac{ca^{bc}}{\Gamma(b)} \left(\frac{\Gamma\left(b+\frac{1}{c}\right)}{a\Gamma\left(b\right)}\right)^2 \frac{\Gamma(b)}{a^{bc}c} \\
&= \frac{\Gamma\left(b+\frac{2}{c}\right)}{a^2\Gamma(b)} - \frac{\left(\Gamma\left(b+\frac{1}{c}\right)\right)^2}{a^2\left(\Gamma(b)\right)^2}.
\end{aligned}$$

The fourth central moment (kurtosis) is

$$\int_0^\infty (x-\mu)^4 f_x dx = \int_0^\infty (x-\mu)^4 \frac{ac\,(ax)^{bc-1}\exp\left(-(ax)^c\right)}{\Gamma(b)}dx$$
$$= \frac{ca^{bc}}{\Gamma(b)}\int_0^\infty (x-\mu)^4 x^{bc-1}\exp\left(-(ax)^c\right)dx$$
$$= \frac{ca^{bc}}{\Gamma(b)}\int_0^\infty \left(x^{bc+3} - 4x^{bc+2}\mu + 6x^{bc+1}\mu^2 - 4x^{bc}\mu^3 + x^{bc-1}\mu^4\right)\exp\left(-(ax)^c\right)dx$$
$$= \frac{1}{a^4}\left(\frac{\Gamma\left(b+\frac{4}{c}\right)}{\Gamma(b)} - 4\frac{\Gamma\left(b+\frac{1}{c}\right)\Gamma\left(b+\frac{3}{c}\right)}{(\Gamma(b))^2} + 6\frac{\left(\Gamma\left(b+\frac{1}{c}\right)\right)^2\Gamma\left(b+\frac{2}{c}\right)}{(\Gamma(b))^3}\right.$$
$$\left. -3\frac{\left(\Gamma\left(b+\frac{1}{c}\right)\right)^4}{(\Gamma(b))^4}\right).$$

The terms of the sum are integrated the same way as in the second central moment integral.

A.1.2 Laplace Distribution

The Laplace distribution is sometimes referred to as the double exponential distribution. Let x follow the Laplace distribution with a probability density function f_x,

$$f_x = \frac{\lambda}{2}\exp\left(-\lambda\left|x-\mu\right|\right),$$

where μ is a location parameter and λ is a real positive number. The first moment (expected value) is

$$\int_{-\infty}^\infty x f_x dx = \int_{-\infty}^\infty x\frac{\lambda}{2}\exp\left(-\lambda\left|x-\mu\right|\right)dx = \mu.$$

Let n be a positive integer. The nth central moment is

$$
\begin{aligned}
\mu_n &= \int_{-\infty}^{\infty} (x-\mu)^n f_x dx \\
&= \frac{\lambda}{2}\left(\int_{-\infty}^{\mu} (x-\mu)^n \exp\left(-\lambda\left(\mu - x\right)\right) dx + \int_{\mu}^{\infty} (x-\mu)^n \exp\left(-\lambda\left(x-\mu\right)\right) dx\right) \\
&= \frac{\lambda}{2}\left(\int_{-\infty}^{0} \left(\frac{y}{\lambda}\right)^n \exp\left(y\right)\frac{dy}{\lambda} + \int_{0}^{\infty} \left(\frac{y}{\lambda}\right)^n \exp\left(-y\right)\frac{dy}{\lambda}\right) \qquad (y = \lambda\left(x-\mu\right)) \\
&= \frac{1}{2\lambda^n}\left(\int_{-\infty}^{0} y^n \exp\left(y\right) dy + \int_{0}^{\infty} y^n \exp\left(-y\right)\frac{dy}{\lambda}\right) \\
&= \frac{1}{2\lambda^n}\left((-1)^n \Gamma(n+1) + \Gamma(n+1)\right) = (-1)^n \frac{n!}{2\lambda^n} + \frac{n!}{2\lambda^n},
\end{aligned}
$$

since $\Gamma(n+1) = n!$, where $\Gamma(\cdot)$ is the Euler's Gamma function. Since Laplace distribution is symmetric, all the odd moments are zero. The nth central moment is then

$$
\mu_n = \begin{cases} 0 & \text{if } n \text{ odd}, \\ \frac{n!}{\lambda^n} & \text{if } n \text{ even}. \end{cases}
$$

The second (variance) and the fourth (kurtosis) central moments are

$$
\begin{aligned}
\mu_2 &= \frac{2}{\lambda^2}, \\
\mu_4 &= \frac{24}{\lambda^4}.
\end{aligned}
$$

A.1.3 Normal Distribution

Let x be normally distributed with a probability density function f_x,

$$
f_x = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right),
$$

where μ is a location parameter and σ is the standard deviation. For $\mu = 0$ and $\sigma = 1$ we refer to this distribution as the standard normal distribution. The first moment (expected value) is

$$
\int_{-\infty}^{\infty} x f_x dx = \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx = \mu.
$$

Let n be a positive integer. The nth central moment is

$$\begin{aligned}
\mu_n &= \int_{-\infty}^{\infty} (x-\mu)^n f_x dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} (x-\mu)^n \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx && (y = x-\mu) \\
&= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} y^n \exp\left(-\frac{1}{2}\left(\frac{y}{\sigma}\right)^2\right) dy && (y = \sigma u) \\
&= \int_{-\infty}^{\infty} \frac{\sigma^n}{\sigma\sqrt{2\pi}} u^n \exp\left(-\frac{u^2}{2}\right) \sigma du = \sigma^n \int_{-\infty}^{\infty} \frac{u^n}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du
\end{aligned}$$

Note that the integral is zero for odd n since the integral of an odd function over the real line is zero. For an even n we can write the integral as

$$\begin{aligned}
&2\sigma^n \int_0^{\infty} \frac{u^n}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du \qquad \left(u = \sqrt{2w} \text{ and } n = 2n\right) \\
&= \sigma^{2n} \frac{2}{\sqrt{2\pi}} \int_0^{\infty} (2w)^n \exp(-w) \frac{dw}{\sqrt{2w}} = \frac{2^n}{\sqrt{\pi}} \int_0^{\infty} w^{n-\frac{1}{2}} \exp(-w)\, dw \\
&= \sigma^{2n} \frac{2^n}{\sqrt{\pi}} \Gamma\left(n+\frac{1}{2}\right) = \sigma^{2n} (2n-1)!!,
\end{aligned}$$

where $\Gamma(\cdot)$ is the Euler's Gamma function and $(\cdot)!!$ is the double factorial. For the Gamma function it holds that

$$\Gamma\left(n+\frac{1}{2}\right) = \frac{(2n-1)!!}{2^n}\sqrt{\pi},$$

see Arfken et al. (2011) for the proof. We get the nth central moment when we substitute n back to the place of $2n$.

$$\mu_n = \begin{cases} 0 & \text{if } n \text{ odd}, \\ \sigma^n (n-1)!! & \text{if } n \text{ even}. \end{cases}$$

The second (variance) and the fourth (kurtosis) central moments are then

$$\begin{aligned}
\mu_2 &= \sigma^2, \\
\mu_4 &= 3\sigma^4.
\end{aligned}$$

A.1.4 Uniform Distribution

Let x be uniformly distributed between finite a and b with a probability density function f_x,

$$f_x = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a,b], \\ 0 & \text{otherwise}. \end{cases}$$

The first moment (expected value) is then

$$\mu = \int_{-\infty}^{\infty} x f_x dx = \int_a^b x f_x dx = \int_a^b x \frac{1}{b-a} dx = \left[\frac{x^2}{2(b-a)}\right]_a^b = \frac{a+b}{2}.$$

Let n be a positive integer. The nth central moment is

$$\begin{aligned}
\mu_n &= \int_{-\infty}^{\infty} (x-\mu)^n f_x dx = \int_a^b (x-\mu)^n \frac{1}{b-a} dx = \left[\frac{(x-\mu)^{n+1}}{(n+1)(b-a)}\right]_a^b \\
&= \frac{(b-\mu)^{n+1} - (a-\mu)^{n+1}}{(n+1)(b-a)} = \frac{\left(\frac{1}{2}(b-a)\right)^{n+1} - \left(\frac{1}{2}(a-b)\right)^{n+1}}{(n+1)(b-a)} \\
&= \frac{(b-a)^n + (a-b)^n}{2^{n+1}(n+1)} = \begin{cases} 0 & \text{if } n \text{ odd}, \\ \frac{(b-a)^n}{2^n(n+1)} & \text{if } n \text{ even}. \end{cases}
\end{aligned}$$

Since the uniform distribution is symmetric, all the odd central moments are 0. The second central moment (variance) and the fourth central moment (kurtosis) are then

$$\begin{aligned}
\mu_2 &= \frac{(b-a)^2}{12}, \\
\mu_4 &= \frac{(b-a)^4}{80}.
\end{aligned}$$

A.2 Complex $\mathrm{Cov}_4(\cdot)$ Scaling

The complex version of the $\mathrm{Cov}_4(\cdot)$ scatter matrix functional is defined the following way

$$\begin{aligned}
&\mathrm{Cov}_4(F_x) = S_2(F_x) = \\
&\frac{1}{p+1}\mathbb{E}\left((x-\mathbb{E}(x))(x-\mathbb{E}(x))^* S_1(F_x)^{-1}(x-\mathbb{E}(x))(x-\mathbb{E}(x))^*\right),
\end{aligned}$$

where $S_1(F_x)$ is the covariance matrix. Let a random vector z with a cdf F_z follow the complex normal distribution with parameters μ, Σ and C,

$$\begin{aligned}
&\mu = \mathbb{E}(z) = 0, \\
&\Sigma = \mathrm{Cov}(F_z) = \mathbb{E}(zz^*) = I_p, \\
&C = \mathbb{E}\left(zz^T\right) = 0_p.
\end{aligned}$$

The complex normal distribution with these parameters is referred to as the standard complex normal distribution. We can present the standard complex normal distribution as $2p$ real normal distribution. If $z = x + yi$, then

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim \mathcal{N}\left(0_{2p}, \frac{1}{2} I_{2p}\right).$$

Hereby, the second and fourth central moments of x_j and y_j are $\frac{1}{2}$ and $\frac{3}{4}$ respectively. The $\mathrm{Cov}_4(\cdot)$ scatter matrix for the standard complex normal distribution is then:

$$\begin{aligned}
&\mathrm{Cov}_4(F_z) = \frac{1}{p+1}\mathbb{E}\left(zz^* I_p zz^*\right) \\
&= \frac{1}{p+1}\mathbb{E}\left(\begin{pmatrix} z_1 \\ \vdots \\ z_p \end{pmatrix}\begin{pmatrix} \bar{z}_1 & \dots & \bar{z}_p \end{pmatrix}\begin{pmatrix} z_1 \\ \vdots \\ z_p \end{pmatrix}\begin{pmatrix} z_1 & \dots & z_p \end{pmatrix}\right) \\
&= \frac{1}{p+1}\mathbb{E}\left(\begin{pmatrix} |z_1|^4 + |z_1|^2 \sum_{j\neq 1}^{p} |z_j|^2 & \dots & \cdot \\ \vdots & \ddots & \vdots \\ \cdot & \dots & |z_p|^4 + |z_p|^2 \sum_{j\neq p}^{p} |z_j|^2 \end{pmatrix}\right) \\
&= \frac{1}{p+1}\mathbb{E}\left(\begin{pmatrix} x_1^4 + 2x_1^2 y_1^2 + y_1^4 + (x_1^2 + y_1^2)\sum_{j\neq 1}^{p}\left(x_j^2 + y_j^2\right) & \dots & \cdot \\ \vdots & \ddots & \vdots \\ \cdot & \dots & \cdot \end{pmatrix}\right) \\
&= \frac{1}{p+1}\begin{pmatrix} \frac{3}{4} + \frac{2}{4} + \frac{3}{4} + (p-1) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{3}{4} + \frac{2}{4} + \frac{3}{4} + (p-1) \end{pmatrix} = I_p,
\end{aligned}$$

where $\bar{z}_i$ is the complex conjugate of z_i. Hereby, $\frac{1}{p+1}$ is the correct scaling factor if we want to scale $\mathrm{Cov}_4(\cdot)$ to be the identity matrix for the standard complex normal distribution.

A.3 Cube to Sphere Mapping

Let

$$S^2 = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}$$

be the unit sphere and

$$C = \{(x, y, z) \in \mathbb{R}^3 : x, y, z \in \{1, -1\}\}$$

be the unit cube surface. A mapping from the cube surface to the unit sphere

$$\pi : C \to S$$

can be formulated as:

$$\pi(x,y,z) = \begin{pmatrix} x\sqrt{1-\frac{y^2}{2}-\frac{z^2}{2}+\frac{y^2z^2}{3}}, \\ y\sqrt{1-\frac{z^2}{2}-\frac{x^2}{2}+\frac{z^2x^2}{3}}, \\ z\sqrt{1-\frac{x^2}{2}-\frac{y^2}{2}+\frac{x^2y^2}{3}} \end{pmatrix} = (X,Y,Z),$$

The inverse mapping from sphere to cube is not as convenient as the cube to sphere mapping. Furthermore, the inverse transformation can not be written in closed form. The R code for the mappings are in Appendix C.1.

A.4 Stereographic Projection

Let

$$S^2 = \{(x,y,z) \in \mathbb{R}^3 : x^2+y^2+z^2 = 1\}$$

be the unit sphere and let n denote the north pole (0,0,1). The stereographic projection map π,

$$\pi : S^2 - n \to \mathbb{C},$$

can be formulated the following way:

$$\pi(x,y,z) = \left(\frac{x}{1-z}, \frac{y}{1-z}\right) = (X,Y).$$

The inverse map can be formulated as:

$$\pi^{-1}(X,Y) = \left(\frac{2X}{1+X^2+Y^2}, \frac{2Y}{1+X^2+Y^2}, \frac{-1+X^2+Y^2}{1+X^2+Y^2}\right) = (x,y,z).$$

Note that π is not defined on n and is a bijective transformation on all the other points of the set. Furthermore, stereographic projection is conformal, meaning that it preserves the angels between curves. The R codes for the mappings are in Appendix C.1.

Appendix B

Proofs of Theorems

B.1 Proof of Theorem 3.2.1

Let x be a p variate random vector with a cumulative distribution function F_x and let $X = [x_1, ..., x_n]$, where $x_1, ..., x_n$ are observations from the distribution F_x. Let $S(F_x)$ be a scatter functional. Now

$$\begin{aligned} &S^{-1/2}\left(F_x\right) A^{-1}\left(S\left(F_{Ax+b}\right)\right)\left(S^{-1/2}\left(F_x\right) A^{-1}\right)^* \\ &= S^{-1/2}\left(F_x\right) A^{-1} A S\left(F_x\right) A^*\left(A^*\right)^{-1}\left(S^{-1/2}\left(F_x\right)\right)^* = I_p. \end{aligned}$$

Thus it follows that $S^{-1/2}(F_{Ax+b}) = OS^{-1/2}(F_x)A^{-1}$ for some orthogonal $O = O\left(F_x, A\right)$.

B.2 Proof of Lemma 4.4.1

Let $A = (a_{ij}) = (v_{ij} + w_{ij}i)$ be a $p \times p$ matrix with at least one nonzero element in each row. Let $\mathcal{L}$ denote the set of all nonsingular $p \times p$ diagonal

matrices and let $L = (l_{ij}) = (x_{ij} + y_{ij}i) \in \mathcal{L}$. Now

$$
\begin{aligned}
||LA - I_p||_F^2 &= \sum_{i=1}^{p} |a_{ii}l_{ii} - 1|^2 + \sum_{\substack{i=1 \\ i\neq j}}^{p}\sum_{j=1}^{p} |a_{ij}l_{ij}|^2 \\
&= \sum_{i=1}^{p} |(v_{ii} + w_{ii}i)(x_{ii} + y_{ii}i) - 1|^2 + \sum_{\substack{i=1 \\ i\neq j}}^{p}\sum_{j=1}^{p} |(v_{ij} + w_{ij}i)(x_{ij} + y_{ij}i)| \\
&= \sum_{i=1}^{p} \left((v_{ii}x_{ii} - w_{ii}y_{ii} - 1)^2 + (v_{ii}y_{ii} + w_{ii}x_{ii})^2\right) \\
&+ \sum_{\substack{i=1 \\ i\neq j}}^{p}\sum_{j=1}^{p} \left((v_{ij}x_{ij} - w_{ij}y_{ij})^2 + (v_{ij}y_{ij} + w_{ij}x_{ij})^2\right) \\
&= \sum_{i=1}^{p} \left(v_{ii}^2x_{ii}^2 + v_{ii}^2y_{ii}^2 + w_{ii}^2x_{ii}^2 + w_{ii}^2y_{ii}^2 - 2v_{ii}x_{ii} + 2w_{ii}y_{ii} + 1\right) \\
&+ \sum_{\substack{i=1 \\ i\neq j}}^{p}\sum_{j=1}^{p} \left(v_{ij}^2x_{ij}^2 + v_{ij}^2y_{ij}^2 + w_{ij}^2x_{ij}^2 + w_{ij}^2y_{ij}^2\right) \\
&= \sum_{i=1}^{p}\sum_{j=1}^{p} \left(x_{ij}^2\left(v_{ij}^2 + w_{ij}^2\right) + y_{ij}^2\left(v_{ij}^2 + w_{ij}^2\right)\right) - 2\sum_{i=1}^{p}(v_{ii}x_{ii} - w_{ii}y_{ii}) + p.
\end{aligned}
$$

Then

$$
\frac{\partial}{\partial x_{ii}}||LA - I_p||_F^2 = 2\left(\sum_{j=1}^{p}\left(x_{ii}\left(v_{ij}^2 + w_{ij}^2\right)\right) - 2v_{ii}\right)
$$

and

$$
\frac{\partial}{\partial y_{ii}}||LA - I_p||_F^2 = 2\left(\sum_{j=1}^{p}\left(y_{ii}\left(v_{ij}^2 + w_{ij}^2\right)\right) + 2w_{ii}\right).
$$

The derivatives are zero when

$$
x_{ii} = \frac{v_{ii}}{\sum_{j=1}^{p}\left(v_{ij}^2 + w_{ij}^2\right)} = \frac{\mathrm{Re}(a_{ii})}{\sum_{j=1}^{p}|a_{ij}|^2}
$$

and

$$
y_{ii} = -\frac{w_{ii}}{\sum_{j=1}^{p}\left(v_{ij}^2 + w_{ij}^2\right)} = \frac{-\mathrm{Im}(a_{ii})}{\sum_{j=1}^{p}|a_{ij}|^2}.
$$

The value of $||LA - I_p||_F^2$ is then

$$p - \sum_{i=1}^{p} \frac{|a_{ii}|^2}{\sum_{j=1}^{p} |a_{ij}|^2}$$

Let $\mathcal{P}$ denote the set of all $p \times p$ permutation matrices. Now denote $\hat{G} = \hat{\Gamma}\Omega$ and $\tilde{G_{ij}} = |\hat{G}|_{ij}^2 / \sum_{k=1}^{p} |\hat{G}|_{ik}^2$, $i, j = 1, \ldots, p$. We can then write the minimum distance index as

$$\hat{D} = D(\hat{G}) = \frac{1}{\sqrt{p-1}} \left(p - \max_{P \in \mathcal{P}} \left(\mathrm{tr} \left(P\tilde{G} \right) \right) \right)^{1/2}.$$

Note that for the Frobenius norm has the following property

$$||A||_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2} = \sqrt{\mathrm{tr}(AA^*)},$$

where A^* is the conjugate transpose of A.

Appendix C

R Codes

This Appendix contains the required R codes to reproduce the results of this thesis.

C.1 Fun_Col_Sep.R

```
plot_jpeg = function(jpg, add=FALSE)
{
  res = dim(jpg)[1:2] # get the resolution
  if (!add) # initialize an empty plot area if add==FALSE
    plot(0,0,xlim=c(0,res[1]),ylim=c(0,res[2]),asp=1,type='n',xaxs='i'
         ,yaxs='i',xaxt='n',yaxt='n',xlab='',ylab='',bty='n')
  rasterImage(jpg,0,0,res[1],res[2])
}

#Project the pixels to the surface of the RBG cube
cor_pix <- function(A){
  d1 <- dim(A)[1]
  d2 <- dim(A)[2]
  d3 <- dim(A)[3]

  for( i in 1:d1){
    for( j in 1:d2){
      tmp_max <- max(A[i,j,])
      tmp_min <- min(A[i,j,])
      diff1 <- 1 - tmp_max

      if( diff1 <= tmp_min){
        A[i,j,match(tmp_max,A[i,j,])] <- 1
      }
      else{
        A[i,j,match(tmp_min,A[i,j,])] <- 0
```

```
      }
    }
  }
  return(A)

}

cube_to_sphere <- function(A){
  D <- A
  n <- dim(A)[1]
  p <- dim(A)[2]
  for(i in 1:n){
    for(j in 1:p){
      x <- A[i,j,1]
      y <- A[i,j,2]
      z <- A[i,j,3]
      D[i,j,1] <- x*sqrt(1- y^2/2- z^2/2 + (y^2*z^2)/3)
      D[i,j,2] <- y*sqrt(1- z^2/2- x^2/2 + (x^2*z^2)/3)
      D[i,j,3] <- z*sqrt(1- x^2/2- y^2/2 + (x^2*y^2)/3)
    }
  }
  return(D)
}

sphere_to_cube <- function(A){
  n <- dim(A)[1]
  p <- dim(A)[2]
  D <- array(dim = c(n,p,3))
  for(i in 1:n){
    for(j in 1:p){
      x <- A[i,j,1]
      y <- A[i,j,2]
      z <- A[i,j,3]

      fx = abs(x)
      fy = abs(y)
      fz = abs(z)

      if (fy >= fx && fy >= fz) {
        a2 = x * x * 2.0
        b2 = z * z * 2.0
        inner = -a2 + b2 -3
        innersqrt = -sqrt((inner * inner) - 12.0 * a2)

        if(x == 0.0 || x == -0.0) {
          position.x = 0.0
        }
        else {
          position.x = sqrt(innersqrt + a2 - b2 + 3.0) * 1/sqrt(2)
```

```
  }

  if(z == 0.0 || z == -0.0) {
    position.z = 0.0
  }
  else {
    position.z = sqrt(innersqrt - a2 + b2 + 3.0) * 1/sqrt(2)
  }

  if(position.x > 1.0) position.x = 1.0
  if(position.z > 1.0) position.z = 1.0

  if(x < 0) position.x = -position.x
  if(z < 0) position.z = -position.z

  if (y > 0) {
    #top face
    position.y = 1.0;
  }
  else {
    # bottom face
    position.y = -1.0;
  }
}else if (fx >= fy && fx >= fz) {
  a2 = y * y * 2.0;
  b2 = z * z * 2.0;
  inner = -a2 + b2 -3;
  innersqrt = -sqrt((inner * inner) - 12.0 * a2);

  if(y == 0.0 || y == -0.0) {
    position.y = 0.0;
  }
  else {
    position.y = sqrt(innersqrt + a2 - b2 + 3.0) * 1/sqrt(2)
  }

  if(z == 0.0 || z == -0.0) {
    position.z = 0.0;
  }
  else {
    position.z = sqrt(innersqrt - a2 + b2 + 3.0) * 1/sqrt(2)
  }

  if(position.y > 1.0) position.y = 1.0;
  if(position.z > 1.0) position.z = 1.0;

  if(y < 0) position.y = -position.y;
  if(z < 0) position.z = -position.z;
```

```
      if (x > 0) {
        # right face
        position.x = 1.0;
      }
      else {
        # left face
        position.x = -1.0;
      }
    }else {
      a2 = x * x * 2.0;
      b2 = y * y * 2.0;
      inner = -a2 + b2 -3;
      innersqrt = -sqrt((inner * inner) - 12.0 * a2);

      if(x == 0.0 || x == -0.0) {
        position.x = 0.0;
      }
      else {
        position.x = sqrt(innersqrt + a2 - b2 + 3.0) * 1/sqrt(2)
      }

      if(y == 0.0 || y == -0.0) {
        position.y = 0.0;
      }
      else {
        position.y = sqrt(innersqrt - a2 + b2 + 3.0) * 1/sqrt(2)
      }

      if(position.x > 1.0) position.x = 1.0;
      if(position.y > 1.0) position.y = 1.0;

      if(x < 0) position.x = -position.x;
      if(y < 0) position.y = -position.y;

      if (z > 0) {
        # front face
        position.z = 1.0;
      }
      else {
        # back face
        position.z = -1.0;
      }
    }
    D[i,j,1] <- position.x
    D[i,j,2] <- position.y
    D[i,j,3] <- position.z

  }
}
```

```
  D.scaled <- 1/2*(1+D)
  return(D.scaled)
}
```

```
comptopix <- function(A,he,wi){
  height <- he
  width  <- wi
  TM <- array(dim=c(he,wi,3))
  for(i in 1:height){
    for(j in 1:width){
      a <- Re(A[i,j])
      b <- Im(A[i,j])
      TM[i,j,1] <- (2*a)/(a^2+b^2+1)
      TM[i,j,2] <- (2*b)/(a^2+b^2+1)
      TM[i,j,3] <- (a^2+b^2-1)/(a^2+b^2+1)
    }
  }
  return(TM)
}
```

```
pixtocomp <- function(D){
  #Scale and shift the location to fit the transformation
  A <- D*2 -1
  height <- dim(A)[1]
  width  <- dim(A)[2]
  TM <- matrix(NA,nrow=height,ncol=width)
  for(i in 1:height){
    for(j in 1:width){
      repart <- A[i,j,1]/(1-A[i,j,3])
      impart <- A[i,j,2]/(1-A[i,j,3])
      TM[i,j] <- complex(re=repart,im=impart)
    }
  }
  return(TM)
}

iseven <- function(a){a\%\%2==0}

snakevector <- function(A){
  height <- dim(A)[1]
  width  <- dim(A)[2]
  vec <- c(rep(NA,height*width))
  for(i in 1:width){
    for(j in 1:height){
```

```
      if(!iseven(i)){
        vec[(i-1)*height + j] <- A[j,i]
      }
      if(iseven(i)){
        vec[(i-1)*height + j] <- A[(height+1-j),i]
      }
    }
  }
  return(vec)
}

reversesnake <- function(vec,he,wi){
  height <- he
  width  <- wi
  M <- matrix(NA,ncol=width,nrow=height)
  for(i in 1:width){
    for(j in 1:height){
      if(!iseven(i)){
        M[j,i] <- vec[(i-1)*height + j]
      }
      if(iseven(i)){
        M[(height+1-j),i] <- vec[(i-1)*height + j]
      }
    }
  }
  return(M)
}
```

C.2 ImageSourceSeparation.R

```
library(clue)
library(forecast)
library(mvtnorm)
library(pixmap)
library(jpeg)
# Make sure you have the all the .R files in the same folder
source("Fun_Col_Sep.R")
source("Reqfun.R")

#Run Fractals.R to generate the following figures.
fig1 <- readJPEG("pic1.jpg")
fig2 <- readJPEG("pic2.jpg")
fig3 <- readJPEG("pic3.jpg")

#The function is designed for a 3 figure setting.
#parameters = figure 1, figure 2, figure 3,
#regular vectorization = TRUE, set of taus,
```

```
#name for the figures
ImageSep <- function(F1,F2,F3,Vec,taus,name){
  he <- min(dim(F1)[1],dim(F2)[1],dim(F3)[1])
  wi <- min(dim(F1)[2],dim(F2)[2],dim(F3)[2])

  F1.scale <- F1[1:he,26:(wi+25),]
  F2.scale <- F2[1:he,1:wi,]
  F3.scale <- F3[1:he,1:wi,]

  # Project the colors to the surface of the RBG cube
  F1.corrected <- cor_pix(F1.scale)
  F2.corrected <- cor_pix(F2.scale)
  F3.corrected <- cor_pix(F3.scale)

  nameoriginal <- paste("original",name,".png",sep="")

  # Plot the corrected figures
  png(nameoriginal,width=480,height=150)
  par(mfrow = c(1,3),mar = c(0.1,0.1,0.1,0.1))
  plot_jpeg(F1.corrected)
  plot_jpeg(F2.corrected)
  plot_jpeg(F3.corrected)
  dev.off()

  # Transform the cube surface to sphere surface and
  # stereohraphic transformation from the sphere to the complex plane
  F1.complex <- pixtocomp(cube_to_sphere(F1.corrected))
  F2.complex <- pixtocomp(cube_to_sphere(F2.corrected))
  F3.complex <- pixtocomp(cube_to_sphere(F3.corrected))

  # Choose the vectorization
  if(Vec ==TRUE){
    Z <- cbind(as.vector(F1.complex),as.vector(F2.complex),
 as.vector(F3.complex))
  }
  else{
    Z <- cbind(snakevector(F1.complex),snakevector(F2.complex),
  snakevector(F3.complex))
  }

  #Choose the mixing matrix
  a <- complex(re=0.1, im =0.04)
  b <- complex(re=0.3, im =0.051)
  c <- complex(re=0.38, im =0.01)
  d <- complex(re=0.12, im =0.078)
  e <- complex(re=0.51, im =0.34)
  f <- complex(re=0.34, im =0.47)
  g <- complex(re=0.01, im =0.032)
  h <- complex(re=0.08, im =0.012)
```

```
i <- complex(re=0.2, im =0.21)

OMEGA <- matrix(c(a,b,c,d,e,f,g,h,i),3,3)

MIXED <- Z \%\*\% t(OMEGA)

M1 <- matrix(MIXED[,1],he,wi)
M2 <- matrix(MIXED[,2],he,wi)
M3 <- matrix(MIXED[,3],he,wi)

#Transform from complex plane to sphere and from
#sphere to cube surface
if(Vec ==TRUE){
  M1.RBG <- sphere_to_cube(comptopix(M1,he,wi))
  M2.RBG <- sphere_to_cube(comptopix(M2,he,wi))
  M3.RBG <- sphere_to_cube(comptopix(M3,he,wi))
}
else{
 #Note that for Vec=FALSE the mixed images do not print correctly
  M1.RBG <- sphere_to_cube(comptopix(M1,he,wi))
  M2.RBG <- sphere_to_cube(comptopix(M2,he,wi))
  M3.RBG <- sphere_to_cube(comptopix(M3,he,wi))
}

namemixed <- paste("mixed",name,".png",sep="")

png(namemixed,width=480,height=150)
par(mfrow = c(1,3),mar = c(0.1,0.1,0.1,0.1))
plot_jpeg(M1.RBG)
plot_jpeg(M2.RBG)
plot_jpeg(M3.RBG)
dev.off()

FOBI.EST <- gammaFOBI(MIXED)
FOBI.GAMMA <- gammaFOBI2(MIXED)
tmp.FOBI <- MD_fun(FOBI.GAMMA,OMEGA)
print(tmp.FOBI)

if(Vec ==TRUE){
  EST1.FOBI <- matrix(FOBI.EST[,1],he,wi)
  EST2.FOBI <- matrix(FOBI.EST[,2],he,wi)
  EST3.FOBI <- matrix(FOBI.EST[,3],he,wi)
}
else{
  EST1.FOBI <- reversesnake(matrix(FOBI.EST[,1],he,wi),he,wi)
  EST2.FOBI <- reversesnake(matrix(FOBI.EST[,2],he,wi),he,wi)
  EST3.FOBI <- reversesnake(matrix(FOBI.EST[,3],he,wi),he,wi)
}
```

```
  EST1.FOBI.RBG <- sphere_to_cube(comptopix(EST1.FOBI,he,wi))
  EST2.FOBI.RBG <- sphere_to_cube(comptopix(EST2.FOBI,he,wi))
  EST3.FOBI.RBG <- sphere_to_cube(comptopix(EST3.FOBI,he,wi))

  namefobi <- paste("fobiunmixed",name,".png",sep="")

  png(namefobi,width=480,height=150)
  par(mfrow = c(1,3),mar = c(0.1,0.1,0.1,0.1))
  plot_jpeg(EST1.FOBI.RBG)
  plot_jpeg(EST2.FOBI.RBG)
  plot_jpeg(EST3.FOBI.RBG)
  dev.off()
  n <- length(taus)

  for(i in 1:1){
    AMUSE.EST <- gammaAMUSE(MIXED,taus[i])
    EST.AMUSE.MAT <- gammaAMUSE2(MIXED,taus[i])
    tmp.AMUSE <- MD_fun(EST.AMUSE.MAT,t(OMEGA))
    print(i)
    print(tmp.AMUSE)

    if(Vec == TRUE){
      EST1.AMUSE <- matrix(AMUSE.EST[,1],he,wi)
      EST2.AMUSE <- matrix(AMUSE.EST[,2],he,wi)
      EST3.AMUSE <- matrix(AMUSE.EST[,3],he,wi)
    }
    else{
      EST1.AMUSE <- reversesnake(matrix(AMUSE.EST[,1],he,wi),he,wi)
      EST2.AMUSE <- reversesnake(matrix(AMUSE.EST[,2],he,wi),he,wi)
      EST3.AMUSE <- reversesnake(matrix(AMUSE.EST[,3],he,wi),he,wi)
    }

    EST1.AMUSE.RBG <- sphere_to_cube(comptopix(EST1.AMUSE,he,wi))
    EST2.AMUSE.RBG <- sphere_to_cube(comptopix(EST2.AMUSE,he,wi))
    EST3.AMUSE.RBG <- sphere_to_cube(comptopix(EST3.AMUSE,he,wi))

    picname <- paste("amuseunmixed_tau",taus[i],name,".png",sep="")
    png(picname,width=480,height=150)
    par(mfrow = c(1,3),mar = c(0.1,0.1,0.1,0.1))
    plot_jpeg(EST1.AMUSE.RBG)
    plot_jpeg(EST2.AMUSE.RBG)
    plot_jpeg(EST3.AMUSE.RBG)
    dev.off()
  }
}
ImageSep(fig1,fig2,fig3,TRUE,c(1,5,10,25),"vec_30_1_2016")
ImageSep(fig1,fig2,fig3,FALSE,c(1,5,10,25),"vecs_30_1_2016")
```

C.3 MDI.R

```
library(clue)
library(forecast)
library(mvtnorm)
source("Reqfun.R")
pMatrix.min <- function(A){
  cost <- t(apply(A^2, 1, sum) - 2 * A + 1)

  vec <- c(solve_LSAP(cost))


  list(A=A[vec,], pvec=vec)
}

MD_fun <- function(W.hat,A)
{
  G <- W.hat \%\*\% A
  RowNorms <- sqrt(rowSums(abs(G)^2))
  G.0 <- sweep(abs(G),1,RowNorms, "/")
  G.tilde <- G.0^2

  p <- nrow(A)
  Pmin <- pMatrix.min(G.tilde)
  G.tilde.p <- Pmin$A

  md <- sqrt(p - sum(diag(G.tilde.p)))/sqrt(p-1)
  return(md)
}
```

C.4 Reqfun.R

```
msqrt <- function(X){ #matrix squareroot
  #Eigendecomposition
  eigen(X)$vectors \%\*\% (diag(eigen(X)$values)^(1/2) \%\*\%
  Conj(t(eigen(X)$vectors)))
}

msqrt2 <- function(A){
  tmp <- svd(A)
  result <- tmp$u \%\*\% diag((tmp$d)^(1/2)) \%\*\% Conj(t(tmp$v))
  return(result)
}

cov1 <- function(X){ #covariance matrix
```

```
  ave <- apply(X,2,mean)
  cent <- sweep(X,2,ave,"-")
  return(1/(dim(X)[1]-1) * t(Conj(cent)) \%\*\% cent)
}

#Named cov2 since package JADE has a cov4() function
cov2 <- function(X){ #cov4 matrix
  p <- dim(X)[2]
  n <- dim(X)[1]
  ave <- apply(X,2,mean)
  cent <- sweep(X,2,ave,"-")
  S1 <- cov1(X)
  SQ1 <- solve(msqrt(S1))
  Z=matrix(data=NA, nrow=n, ncol=p)
  for(k in 1:n){
    Z[k,] <- SQ1 \%\*\% cent[k,]
  }
  K = matrix(0,nrow=p,ncol=p)
  for(k in 1:n){
    a <- Z[k,] \%\*\% t(Conj(Z[k,])) \%\*\%  Z[k,]
         \%\*\% t(Conj(Z[k,]))
    K <- K + a
  }
  msqrt(S1) \%\*\% K \%\*\% msqrt(S1) /(n*(p+2))
}
acov <- function(X,tau){ #autocovariance matrix
  n <- dim(X)[1]
  ave <- apply(X,2,mean)
  Z <- sweep(X,2,ave,'-')
  tmp1 <-  Conj(t( Z[1:(n-tau),] )) \%\*\% Z[(1+tau):n,]
  tmp2 <-  Conj(t( Z[(1+tau):n,] )) \%\*\% Z[1:(n-tau),]
  AC <- 1/(2*(n-tau)) * (tmp1 + tmp2)
  return(AC)
}
gammaAMUSE2 <- function(D,t)
{#return the gamma-matrix estimate
  n <- dim(D)[1]
  p <- dim(D)[2]
  T1 <- apply(D,2,mean)
  cent <- sweep(D,2,T1,'-')
  S1 <- cov1(D)
   #Cov is affine equivariant, lose less precision
  # when calculated from original
  Z <- cent \%\*\% solve(msqrt(S1)) #whitening
  S2 <- acov(Z,t)
  U2 <- eigen(S2)$vectors
  G <-t(  Conj(t(U2)) \%\*\% t(solve(msqrt(S1)))  )
  #Gamma transpose
  return(G)
```

```
}
gammaAMUSE <- function(D,t)
{#returns the transformed data
  n <- dim(D)[1]
  p <- dim(D)[2]
  T1 <- apply(D,2,mean)
  cent <- sweep(D,2,T1,'-')
  S1 <- cov1(D)
   #Cov is affine equivariant, lose less precision
  # when calculated from original
  Z <- cent \%\*\% solve(msqrt(S1))
  S2 <- acov(Z,t)
  U2 <- eigen(S2)$vectors
  DT <- Z \%\*\% U2
  return(DT)
}
gammaFOBI2 <- function(D){
  #return the gamma matrix estimate
  n <- dim(D)[1]
  p <- dim(D)[2]
  T1 <- apply(D,2,mean)
  cent <- sweep(D,2,T1,'-')
  S1 <- cov1(D)
  #Cov is affine equivariant, lose less precision
  # when calculated from original
  Z <- cent \%\*\% solve(msqrt(S1)) #whitening
  S2 <- cov2(Z)
  U2 <- eigen(S2)$vectors
  G <-t(  Conj(t(U2)) \%*\% t(solve(msqrt(S1)))  )
  #Gamma transpose
  return(G)
}
gammaFOBI <- function(D){
  #returns the transformed data
  n <- dim(D)[1]
  p <- dim(D)[2]
  T1 <- apply(D,2,mean)
  cent <- sweep(D,2,T1,'-')
  S1 <- cov1(D)
  #Cov is affine equivariant, lose less precision
  #when calculated from original
  Z <- cent \%\*\% solve(msqrt(S1))
  S2 <- cov2(Z)
  U2 <- eigen(S2)$vectors
  DT <- Z \%\*\% U2
  return(DT) #Gamma transpose
}
```

Bibliography

Adali, T. and V. D. Calhoun (2007), Complex ICA of Brain Imaging Data. *IEEE Signal Processing Magazine*, September, 136–139.

Adali, T., P. J. Schreier, and L. L. Scharf (2011), Complex-Valued Signal Processing: The Proper Way to Deal With Impropriety. *IEEE Transactions on signal processing*, 59, 5101–5125.

Anemüller, J., T. J. Sejnowski, and S. Makeig (2003), Complex independent component analysis of frequency-domain electroencephalographic data. *Neural Networks*, 16, 1311–1323.

Arfken, G. B., H. J. Weber, and F. E. Harris (2011), *Mathematical methods for physicists: A comprehensive guide*. Academic press.

Bar-Ness, Y., J. W. Carlin, and M. L. Steinberger (1982), Bootstrapping adaptive interference cancelers-some practical limitations. In *Globecom'82-Global Telecommunications Conference*, volume 1, 1251–1255.

Brockwell, P. J. and R. A. Davis (2013), *Time series: theory and methods*. Springer Science & Business Media.

Bullmore, E., M. Brammer, S. C. Williams, S. Rabe-Hesketh, N. Janot, A. David, J. Mellers, R. Howard, and P. Sham (1996), Statistical methods of estimation and inference for functional MR image analysis. *Magnetic Resonance in Medicine*, 35, 261–277.

Cardoso, J.-F. (1989), Source separation using higher order moments. In *Acoustics, Speech, and Signal Processing, 1989. ICASSP-89., 1989 International Conference on*, 2109–2112, IEEE.

Chen, A. and P. J. Bickel (2006), Efficient independent component analysis. *The Annals of Statistics*, 34, 2825–2855.

Comon, P. (1994), Independent component analysis, a new concept? *Signal processing*, 36, 287–314.

Comon, P. and C. Jutten (2010), *Handbook of Blind Source Separation: Independent component analysis and applications*. Academic press.

Davies, P. L. (1987), Asymptotic behavior of S-estimates of multivariate location parameters and dispersion matrices. *The Annals of Statistics*, 15, 1269–1292.

Friston, K. J., A. P. Holmes, K. J. Worsley, J.-P. Poline, C. D. Frith, and R. S. Frackowiak (1994), Statistical parametric maps in functional imaging: a general linear approach. *Human brain mapping*, 2, 189–210.

Hogg, R. V., J. W. McKean, and A. T. Craig (2005), *Introduction to Mathematical Statistics*. Prentice Hall.

Horn, R. A. and C. R. Johnson (1985), *Matrix Analysis*. Cambridge University Press, New York.

Hwang, S. and S. E. Satchell (1999), Modelling emerging market risk premia using higher moments. *Return Distributions in Finance*, 75.

Hyvärinen, A., J. Karhunen, and E. Oja (2001), *Independent Component Analysis*. John Wiley & Sons, New York.

Ilmonen, P. (2013), On asymptotic properties of the scatter matrix based estimates for complex valued independent component analysis. *Statistics & Probability Letters*, 83, 1219–1226.

Ilmonen, P., J. Nevalainen, and H. Oja (2010a), Characteristics of multivariate distributions and the invariant coordinate system. *Statistics & probability letters*, 80, 1844–1853.

Ilmonen, P., K. Nordhausen, H. Oja, and E. Ollila (2010b), A new performance index for ICA: properties, computation and asymptotic analysis. In *Latent Variable Analysis and Signal Separation*, 229–236, Springer.

Ilmonen, P., K. Nordhausen, H. Oja, and E. Ollila (2012a), On asymptotics of ICA estimators and their performance indices. *arXiv preprint arXiv:1212.3953*.

Ilmonen, P., H. Oja, and R. Serfling (2012b), On invariant coordinate system (ICS) functionals. *International Statistical Review*, 80, 93–110.

Ilmonen, P., D. Paindaveine, et al. (2011), Semiparametrically efficient inference based on signed ranks in symmetric independent component models. *the Annals of Statistics*, 39, 2448–2476.

Kent, J. T., D. E. Tyler, et al. (1996), Constrained M-estimation for multivariate location and scatter. *The Annals of Statistics*, 24, 1346–1370.

Kiviluoto, K. and E. Oja (1998), Independent Component Analysis for Parallel Financial Time Series. In *ICONIP*, volume 2, 895–898.

Lange, N. (2003), What can modern statistics offer imaging neuroscience? *Statistical methods in medical research*, 12, 447–469.

Lietzén, N., P. Ilmonen, and K. Nordhausen (2016), Complex valued AMUSE – performance and asymptotic analysis. *Unpublished Manuscript*.

Lopuhaä, H. P. (1989), On the relation between S-estimators and M-estimators of multivariate location and covariance. *The Annals of Statistics*, 17, 1662–1683.

Mandic, D., S. Javidi, S. Goh, A. Kuh, and K. Aihara (2009), Complex-valued prediction of wind profile using augmented complex statistics. *Renewable Energy*, 34, 196–201.

Maronna, R. A., R. D. Martin, and V. J. Yohai (2006), Wiley Series in Probability and Statistics. *Robust Statistics: Theory and Methods*, 404–414.

Matteson, D. S. and R. S. Tsay (2011), Dynamic orthogonal components for multivariate time series. *Journal of the American Statistical Association*, 106.

Miettinen, J., K. Nordhausen, H. Oja, and S. Taskinen (2012), Statistical properties of a blind source separation estimator for stationary time series. *Statistics and Probability Letters*, 82, 1865–1873.

Nordhausen, K. (2014), On robustifying some second order blind source separation methods for nonstationary time series. *Statistical Papers*, 55, 141–156.

Nordhausen, K., J. Cardoso, H. Oja, and E. Ollila (2011a), JADE: JADE and ICA performance criteria. *R package version*, 1, 0–4.

Nordhausen, K., P. Ilmonen, A. Mandal, H. Oja, and E. Ollila (2011b), Deflation-based FastICA reloaded. In *Proceedings of 19th European Signal Processing Conference*, 1854–1858.

Nordhausen, K., H. Oja, and E. Ollila (2008), Robust independent component analysis based on two scatter matrices. *Austrian Journal of Statistics*, 37, 91–100.

Nordhausen, K., E. Ollila, and H. Oja (2011c), On the performance indices of ICA and blind source separation. In *Signal Processing Advances in Wireless Communications (SPAWC), 2011 IEEE 12th International Workshop on*, 486–490, IEEE.

Oja, H., S. Sirkiä, and J. Eriksson (2006), Scatter matrices and independent component analysis. *Austrian Journal of Statistics*, 35, 175–189.

Ollila, E. (2010), The deflation-based FastICA estimator: statistical analysis revisited. *Signal Processing, IEEE Transactions on*, 58, 1527–1541.

Ollila, E., H. Oja, and V. Koivunen (2008), Complex-valued ICA based on a pair of generalized covariance matrices. *Computational Statistics & Data Analysis*, 52, 3789–3805.

Paindaveine, D. (2006), A Chernoff–Savage result for shape: On the non-admissibility of pseudo-Gaussian methods. *Journal of Multivariate Analysis*, 97, 2206–2220.

Papadimitriou, C. H. and K. Steiglitz (1982), *Combinatorial optimization: algorithms and complexity*. Courier Corporation.

Roll, J.-P. (1981), *Contribution à la Proprioception Musculaire, à la Perception et au Contrôle du Mouvement Chez l'Homme*. Ph.D. thesis.

Samworth, R. J., M. Yuan, et al. (2012), Independent component analysis via nonparametric maximum likelihood estimation. *The Annals of Statistics*, 40, 2973–3002.

Singh, D. K., S. Tripathi, and P. Kalra (2006), Separation of image mixture using complex ICA. In *The 9th Asian Symposium on Information Display ASID06*, 314–317.

Tong, L., V. Soon, Y. Huang, and R. Liu (1990), AMUSE: a new blind identification algorithm. In *Circuits and Systems, 1990., IEEE International Symposium on*, 1784–1787, IEEE.

Tyler, D. E., F. Critchley, L. Dümbgen, and H. Oja (2009), Invariant coordinate selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71, 549–592.

Welvaert, M., J. Durnez, B. Moerkerke, G. Verdoolaege, and Y. Rosseel (2011), neuRosim: An R package for generating fMRI data. *Journal of Statistical Software*, 44, 1–18.